DOCUMENT RESUME

ED 171 329 IR 007 416

AUTHOR Simonson, Michael R., Ed.; Grammer, Bobbie, Ed.
TITLE Proceedings of Selected Research Paper Presentations at the 1979 Convention of the Association for Educational Communications and Technology.
INSTITUTION Association for Educational Communications and Technology, Washington, D.C.
PUB DATE Mar 79
NOTE 738p.; Papers presented at the Research and Theory Division of the Association for Educational Communications and Technology (New Orleans, Louisiana, March 1979) ; Best copy available

EDRS PRICE MF04/PC30 Plus Postage.
DESCRIPTORS *Audiovisual Instruction; Educational Objectives; *Educational Research; *Educational Technology; Evaluation; *Instructional Media; Learning Characteristics; Measurement; *Media Research; Teaching Methods
IDENTIFIERS Association Educational Communications Technology; *Educational Design

ABSTRACT

This collection of 32 research papers presented to the Research and Theory Division at the 1979 convention of the Association for Educational Communications and Technology represents some of the most current thinking in the educational communications and technology field. While each paper was intended to stand alone, they have been grouped in five general categories for ease of use: (1) research reviews, reports, and theoretical discussion; (2) learner characteristics and presentation organization research; (3) media characteristics research; (4) measurement and assessment techniques; and (5) project reports and evaluations. The addresses of the authors are provided for readers who may wish to question them concerning their ideas, methods, or conclusions. This is the first year that the Research and Theory Division has collected selected research papers from these presentations for publication as a proceedings; a list of division officers and board members is included. (Author/RAO)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED171329

PROCEEDINGS OF SELECTED RESEARCH PAPER PRESENTATIONS

at the 1979 Convention of the

Association for Educational Communications and Technology

sponsored by the

Research and Theory Division

March, 1979 in New Orleans, LA

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BEST COPY AVAILABLE

Edited by:

Michael R. Simonson, Ph.D.

Associate Professor of Secondary Education

and

Bobbie Grammer

Teaching Assistant

College of Education

Iowa State University

Ames, Iowa 50011

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Michael Simonson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

IR007416

PREFACE

For the first time in 1979, the Research and Theory Division of the Association for Educational Communications and Technology collected selected research papers from those presented at the national convention in New Orleans, LA for publication in this "Proceedings" document. This publication was printed in limited hardcopy quantity, but is available through the Educational Resources Information Clearinghouse (ERIC) system.

REFEREEING PROCESS: All papers selected for presentation at the Annual Convention and included in this "Proceedings" were subjected to a blind refereeing process. Proposals were submitted by potential presenters to Dr. Bill Winn of the University of Calgary who removed all reference to author. Proposals were then submitted to reviewers for evaluation. Approximately forty percent of the manuscripts submitted for consideration were selected for presentation at the Convention and for publication in these "Proceedings." The manuscripts contained in this document represent some of the most current thinking in the educational communications and technology field.

While each paper was intended to stand alone, they have been grouped according to a general category for ease of use. Individual authors should be questioned concerning their ideas, methods, or conclusions.

M. R. S.

Research and Theory Division

RESEARCH AND THEORY DIVISION OFFICERS AND BOARD MEMBERS

Dean R. Spitzer, President
Senior Educational Specialist
Digital Equipment Corporation
Main Street
Boston, Massachusetts 01505

Dr. William Winn, President-Elect
Dept. of Curriculum & Instruction
University of Calgary
Calgary, Alberta
Canada T2N 1N4

Dr. Frank Dwyer, Newsletter Editor
Mitchell Building
The Pennsylvania State University
University Park, Pennsylvania 16802

Tillman Ragan (1978-1981)
College of Education
University of Oklahoma
Norman, Oklahoma 73069

John M. Keller (1978-1981)
Instructional Technology
School of Education
Syracuse University
Syracuse, New York 13210

Dr. Perrin Parkhurst (1976-1979)
Associate Director, Faculty Development
College of Osteopathic Medicine
Michigan State University
East Lansing, Michigan 48823

William G. Holliday (1978-1979)
Dept. of Curriculum & Instruction
University of Calgary
Calgary, Alberta
Canada T2N 1N4

Louis H. Berry (1978-1979)
Educational Communications & Technology
103 LIS Building
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

Michael R. Simonson (1978-1981)
321 Curtiss Hall
College of Education
Iowa State University
Ames, Iowa 50011

TABLE OF CONTENTS

PART I: Research Reviews, Reports, and Theoretical Discussions

Page

1. Research on Media--Where Do We Go From Here?
by Francis E. Clark and Jay F. Angert ---------------------- 1

2. Research Interpretation: A Justification for Media
by Francis M. Dwyer -- 20

3. Behavioral Objectives: A Critical Review of Theory and Research
by Gary L. Filan and Vernon S. Gerlach -------------------- 28

4. Relating Behavioral Science Research to Practice: What, Why, How
by Dr. Malcolm L. Fleming ----------------------------------- 73

5. Applying the ATI Concept in an Operational Environment
by Perrin E. Parkhurst, Ph.D. ------------------------------ 88

6. In Search of a Better Way to Organize Instruction: The Elaboration Theory
by Charles M. Reigeluth ----------------------------------- 100

7. A Consumer's Guide to Common Flaws in Research
by Dennis M. Roberts -------------------------------------- 132

8. Reanalysis of Research Studies in A.V.C.R.
by Dennis E. Sheriff and John A. Horton ------------------ 140

9. Research for the Practitioner
by Brenda H. White -- 144

PART II: Learner Characteristics and Presentation Organization Research

1. Eye Movement Research and the Interaction Between Television and Child Related Characteristics
by Lois J. Baron --- 158

2. Principal's Perceptions of Actual and Ideal Roles of the School Media Specialist
by Sally Burnell --- 191

3. The Effects of Self Evaluation as an Instructional Feedback Strategy
by Lou M. Carey, Larry Israelite, and Richard F. Schmid --- 212

4. An Empirical Analysis of the Characteristics of Performance Objectives
by Vernon S. Gerlach and Richard E. Schmid --------------- 241

5. The Politics of Instructional Innovation in Higher Education: A Cross Cultural Analysis
by John G. Hedberg, Ph. D. ---------------------------------- 244

6. Effect on Intended and Incidental Learning From the Use of Learning Objectives With An Audiovisual Presentation
by Robert Main -- 267

7. A Comparison: Predicting College Level Academic Success with Tests of Cognitive Style and Cognitive Aptitude
by Dr. Thomas Schwen and Anne K. Bednar ------------------ 303

8. Environmental and Personal Factors Affecting Instructional Development by the Media Professional at the K-12 Level
by Philip Turner, Ed. D. and Nina N. Martin, Ed. D. ------- 321

PART III: Media Characteristics Research

1. The Effects of Picture Type and Picture Location on Comprehension
by Philip J. Brody and Alice Legenza ---------------------- 351

2. Recall and Retrieval from Mapped and Programmed Text
by David H. Jonassen, Ed. D. ------------------------------ 364

3. The Instructional Effectiveness of Integrating Abstract and Realistic Visualization
by John H. Joseph --- 380

4. Efficiency of Students' Achievement Using Black/White and Color Learning and Test Materials
by Richard J. Lamberski and Dennis M. Roberts ------------ 417

5. The Effect of Music and Sound Effects on the Listening Comprehension of Fourth Grade Students
by Raymond E. Mann, Ed. D. -------------------------------- 429

6. Children's Understanding of Implied Motion Cues
by Ronald A. Saiet, Ed. D. -------------------------------- 457

7. Effect of Cartoon Illustrations on the Comprehension and Evaluation of Information Presented in the Print and Audio Mode
by Edward H. Sewell, Jr. ---------------------------------- 496

PART IV: Measurement and Assessment Techniques

1. Measuring Attitudes and Instructional Development: Why and How
by Dr. Michael Simonson ----------------------------------- 521

2. Needs Assessment
by Dean R. Spitzer -- 543

3. Visual Literacy Assessment
by Mary Louise Turner ------------------------------------- 552

4. Field-Testing Instructional Materials
by Bill Winn -- 580

PART V: Project Reports and Evaluations

1. Outcome of the Instructional Systems Development Project
by Robert K. Branson -------------------------------------- 602

2. The Design, Development and Evaluation of an Evaluative Computer Simulation
by Lisa R. Ehrlich, M.A. ---------------------------------- 631

3. The Formative Evaluation of Bilingual Television Programs: Some Results and Suggestions for Improvement
by Richard F. Lewis, Ph.D. -------------------------------- 650

4. Educational Development: The Republic of Korea (1970-1979)
by Robert M. Morgan --------------------------------------- 682

PART I:

Research Reviews, Reports and Theoretical Discussions

Title • RESEARCH ON MEDIA -- WHERE DO WE GO FROM HERE?

Authors • Francis E. Clark, Director
Educational Technology
College of Education
Texas A&M University
College Station, Texas 77843

• Jay F. Angert
Educational Technology
College of Education
Texas A&M University
College Station, Texas 77843

Title • RESEARCH ON MEDIA -- WHERE DO WE GO FROM HERE?

Authors • FRANCIS E. CLARK
Texas A&M University

• JAY F. ANGERT
Texas A&M University

Abstract • A summarization of media research inadequacies remains a standard feature of media literature reviews. A Gestalt solution would be a research model, applicable to field studies in educational institutions, which accounts for the total instructional setting. It appears that an appropriate research model must be directly related to the psychological, sociological, and physiological attributes of the learner, the teacher, the task, and the resources, which collectively result in an instructional environment. The objective form (physiological) and subjective meaning (sociological) of any learning task must yield a functional distinctiveness (psychological) in terms of the sensory information to be extracted by the learner.

RESEARCH ON MEDIA -- WHERE DO WE GO FROM HERE?

FRANCIS E. CLARK
Texas A&M University
JAY F. ANGERT
Texas A&M University

A summarization of media research inadequacies remains a standard feature of literature reviews. A serious reader of the research literature for the past ten years will encounter numerous depressing appraisals of the limited value of media research for improving education. Although many reviewers have decried the uncertain quality and utility of media research, an exhaustive litany of futility is unnecessary. The comments which follow are sufficiently illustrative.

Research Deficiencies

In 1968, Snow and Salomon remarked that "virtually nothing is known . . . about the teaching effectiveness of instructional media" (p. 341). This conclusion was based, in part, on their observation of the widespread use of experimental designs which averaged individual learner differences, although the prime importance of these differences as independent variables had been previously noted (Cronbach, 1957). Comstock (1975) similarly concluded that the utility of media research for either theory or practice was inconsequential. In addition, research with, as opposed to research on media, has been the rule rather than the exception (Salomon, 1970). Researchers have repeatedly treated a given medium as a whole entity, as in comparison studies of film versus television, in an attempt to support the premise that the media could indeed teach. Fleming (1970) recognized that such gross comparisons yielded meaningless data since they masked considerably more variability than they explained. In 1977, Schramm described this macro quality as perhaps the most regrettable feature of the long list of instructional media experiments. Levie and Dickie (1973) suggested that research would be better conceptualized by specifying media variables in terms of specific attributes. Finally, Conway (1970) and Dwyer (1972) have identified as a major research deficiency the lack of logical correlations between the treatment content encountered in many studies and that typical of actual classroom instruction. Glaser (1972) and Salomon and Snow (1970) have also noted the pressing need for "ecological validity," i.e., experimentation under normal instructional conditions.

Perhaps the most brutally frank assessment of media research was offered by Hawkridge (1973):

> The fact is that instructional researchers and designers have not provided even the foundations for constructing strong practical procedures for selecting media appropriate to given learning tasks. If there has been British work in this area, I have been unable to discover it . . . . In the United States, over 2000 media studies have not yielded the answers we need. (p. 1)

Taken at face value, these assessments imply a great deal of misdirected energy, over a lengthy time span, to establish a data base of questionable value.

ATI as a Solution

The widespread recognition that media research had failed to attend to individual learner differences prompted repeated calls for employing the methodology known synonomously as Trait-Treatment Interaction (TTI), or Aptitude-Treatment Interaction (ATI), hereinafter referred to simply as ATI (Allen, 1971; Berliner and Cahen, 1973; Cronbach, 1957; Cronbach and Snow, 1977; DiVesta, 1975; Snow and Salomon, 1968; Virag, 1976). Within this context, aptitudes or traits are defined broadly enough to include the psychological, sociological and physiological characteristics of learners. Cronbach and Snow (1977) suggested that any aspect of an individual which may be useful in predicting instructional responses should be considered an "aptitude." Treatments are defined in a similarly broad fashion so as to include variations among most experimentally manipulable aspects of the teaching or the environment.

Interactions may be defined statistically as regression slopes which depart from parallelism. A disordinal interaction suggests that different treatments are differentially superior for students who are at different levels of a particular trait or aptitude. An ordinal interaction, however, suggests that one treatment retains its superiority over an alternate treatment throughout the range of aptitudes under consideration. However, this superiority is usually more pronounced at one level of the aptitude than at others (Ott, 1977; Cronbach and Snow, 1977).

Adaptations in education to individual differences are neither new nor difficult to find. Considering the obvious extent to which the educational community has accepted instructional and programmatic practices geared to individual differences, the concerted research efforts to locate educationally relevant ATI surficially makes eminent good sense. It is intellectually difficult to deny that ATI's exist. To do so is tantamount to asserting that the instruction which works best for one group of students is therefore best for all students (Cronbach and Snow, 1977).

Paradoxically, it is the firm belief in human individuality and instructional diversity which has so complicated ATI research. As Cronbach (1975) stated, "Once we attend to interactions, we enter a hall of mirrors which extends to infinity . . ." (p. 119). The greatest difficulty ATI researchers have faced is the isolation of those aptitudes and treatment conditions, from an unknown universe of differences, which reliably interact with particular instructional treatments to produce predictable learning outcomes. Considering the immensity of the task, and the relative infancy of the technique, it is hardly surprising that ATI results have been disappointing (Bracht, 1970; Cronbach and Snow, 1977; Dwyer, 1978; Heidt, 1977; Parkhurst, 1975). Undeniably, ATI results have been less than spectacular. The range of aptitudes and treatments is so vast that researchers have had a veritable field day in devising researchable combinations. A search for ATI's calls for no less than a survey of all the ways in which individuals and instructional treatments may differ. These constructs may pair up to form virtually limitless ATI hypotheses (Cronbach and Snow, 1977). The result has been a bewildering array of studies with relatively few threads of commonality.

In the face of such diversity, the temptation is great to use a shotgun approach in searching for ATI's. Salomon (1971) has noted a tendency on the part of many ATI researchers to include extremely large numbers of trait measures in their studies, in the hope of discovering some interactions. Such interactions, even when found, are weak in their explanatory power, having arisen atheoretically and from an inadequate conceptualization of the traits. For the most part, this broad-band exploration approach has not been successful.

Attempts to integrate the fragmentary ATI results have met with only limited success. Allen (1975) concluded that generalizing from the available results was virtually impossible. The similarity of his comments to those of Hawkridge was striking: ". . . there is little definitive evidence from the aptitude-treatment interaction research that points conclusively to the employment of practices that might guide the selection of the more general instructional strategies, much less lead to the design of specific instructional media" (p. 139). Dwyer (1978) and Parkhurst (1975) have also noted the limited usefulness and meaningfulness of ATI research to date. In the summary of what is undoubtedly the seminal work for research on interactions, Cronbach and Snow (1977) concluded that "No Aptitude X Treatment interactions are so well confirmed that they can be used directly as guides to instruction" (p. 492).

Numerous methodological problems have plagued the search for ATI (Cronbach and Snow, 1977). Quite often, however, investigations which have paid close attention to acceptable methodology and data analysis have frequently paid inadequate attention to the more subtle, but equally vital, manner in which the particular constructs chosen as dependent and independent variables complement one another. Thus, a large portion of the media research shortcomings has stemmed from an inadequate conceptualization of pertinent variables. Snow (1970) pointed out the inappropriateness of the majority of constructs in differential psychology for use in ATI research. Cronbach and Snow (1969) previously cited the need for new conceptualizations of traits and treatments. However, both past and present admonishments have largely gone unheeded.

The ATI research literature is so disparate and contradictory that reviewers find themselves in disagreement over its proper interpretation. How is one to make sense out of a body of research which fails to produce interactions where hypothesized, produces interactions in unanticipated and inexplicable fashion, and which may or may not replicate interactions in subsequent studies? Heidt (1977) stated that "To prove a trait-treatment interaction, it is necessary to detect a disordinal interaction . . ." (p. 13). Heidt further concluded that the ATI results are so inconsistent that general summarization is impossible. Berliner and Cahen (1973), however, argued that ordinal interactions are as useful in ATI research as disordinal. Contrasting reviewer techniques have further muddied the waters. Of the ninety studies which Bracht examined, only five were adjudged as giving adequate evidence of ATI, since they produced disordinal interactions. Cronbach and Snow (1977), however, regarded Bracht's criteria as overly stringent and, in a reexamination of several studies dismissed by Bracht as failing to show ATI's, found disordinal interactions.

As different as non-ATI and ATI research are from one another in methodology and philosophy, it is interesting to note that many of the criticisms leveled against non-ATI research are equally applicable to the newer methodology. To the extent that these inadequacies persist, confusion will still reign. In 1970, Shulman warned against research which measured aptitudes with micrometers but environments with divining rods; yet critics are still decrying the unprofitability of using gross aptitude and treatment measures (Anderson, Ball, and Murphy, 1975; Cronbach and Snow, 1977; Dwyer, 1978). According to Dwyer (1978), unrealistic treatment content is still being experimentally varied under artificial pedagogical conditions (Salomon and Clark, 1977). Methodological problems (e.g., inadequate sample sizes and data analyses, and only rare replications), continue to hamper ATI efforts (Berliner and Cahen, 1973; Cronbach and Snow, 1977).

These difficulties have led every reviewer of ATI literature we have encountered to paint a depressingly familiar picture of ineffectiveness.

All, however, have been loathe to suggest abandoning the effort and, surprisingly, have reached somewhat optimistic conclusions on the heels of pessimistic reviews. Perhaps the single most pervasive shortcoming of ATI research efforts that we have detected is the lack of inclusiveness, i.e., collective inattention to the totality of the learning environment. The term Aptitude-Treatment Interaction by itself denotes an overly simplistic two-dimensional conception of learning environments and has perhaps engendered a delimited focus among some researchers. ATI investigations have provided a forum for researchers to promote a spectrum of variables covering learner, teacher, and treatment characteristics; environmental or situational conditions; and a variety of resource characteristics. If research is to proceed systematically toward usable conclusions, some semblance of order must be imposed on the mass of ATI hypotheses. To date, research has been conducted from each researcher's conception of fundamental combinations of attributes. However, we are unaware of a research model which effectively relates these diverse attributes.

The literature is replete with suggestions. Carpenter (1972) called for a blend of media and modes, instructional functions and objectives, content and audience characteristics, and learning environments, while Clark (1975) stressed the relationship between instructional methods, materials, and individual aptitudes. DiVesta (1975) suggested concentrating on cognitive processing variables, whereas Salomon (1976) argued that a presentation's effectiveness depended on a match of mental skills activated by the presentation's code and the learning task requirements. Finally, Burns (1976) suggested a blend of learner, media and environmental variables, while Schramm (1977) pointed out the need for studies of the content of instructional media.

While these varied recommendations have meritorious features, they reflect both an obvious lack of common terminology and a vast disagreement on a comprehensive model from which research on instructional media selection should proceed.

A Practitioner/Researcher Continuum

Since the empirical data have not provided clear research directions, we are faced with a body of literature from which it is difficult to extract "general" principles. Thus, we will need to extrapolate from a broad spectrum of models, paradigms, classification systems, and hierarchies. In this way we may be able to identify the common denominators of the current literature. Any resulting research model would, of necessity, be eclectic in nature.

While it may be true that there is nothing so practical as a good theory, it is also true that much theorizing has little relationship with practicality. Instructional researchers would do well to adopt, adapt, and apply the eclectic instructional practices of "successful" instructors to the design of instructional research. Undeniably, a considerable amount of classroom instruction, devoid of experimental controls or constraints, frequently produces learning of practical significance. One possible explanation may lie in the holistic approach which is characteristic of the "effective" instructor, but which, philosophically, is worlds apart from most research efforts. Intuitively, many instructors manage to derive an optimal blend of personal style, learner and resource characteristics, and task requirements through a consideration of psychological, sociological, and physiological factors.

All too frequently, however, the researcher operates from a narrowly circumscribed perspective. This tunnel vision tendency leads to narrowly conceived research hypotheses. These hypotheses, in turn, engender explanations of data in terms which either sustain or modify the philosophic bent of a particular researcher. Rarely are philosophic lines crossed to acknowledge equally valid or potentially superior explanations of observed learning outcomes. This omission is true both across and within disciplines. The field of educational technology serves as an appropriate example.

We assert that past media research has been philosophically dichotomous, arising either from practitioner or researcher concerns. Becker (1977) described how such artificial distinctions have excluded consideration of many important variables. Not surprisingly, the various media selection models have closely paralleled these media research directions. Mielke (1973) referred to this separation as the distinction between administrative research and basic research, and Clark (1975) extended this distinction to most media taxonomies.

An Eclectic Model for Research and Instruction

To a large extent, commitment to preferred statistical methodology has also dictated research directions. Usually, investigators have dealt with one or two instructional variables at a time, either in a search for main effects or for interactions. To this end, researchers have used regression analysis and other sophisticated statistical techniques to analyze fairly unsophisticated subjective measures of attributes or traits. For these reasons, statistical trends have failed to produce the consistency needed for the development of an instructional model.

Considering the unproductive history of research on instruction, it seems appropriate to step back and take a second look at the diverse research directions which have been, to a great extent, independent of one another. While we recognize that research frequently transcends artificial boundaries, we nevertheless submit that most research on instruction may be categorized loosely into three major areas. We feel that these areas closely parallel the considerations of the effective instructor described earlier.

The first area may be termed functional and/or differential psychological research. This area deals primarily with intellectual abilities, as well as the relationships among stimuli, mediating covert behaviors, and observable overt responses. More precisely, researchers within this domain of research usually begin with a psychological theory and then proceed to validate the logically derived statements, in the form of constructs, through schemes for organizing data for quantitative analysis. The following schemes are usually thought to be synonomous: theory, model, paradigm, analogy, structure, hierarchy, and system. E. L. Thorndike, J. P. Guilford, R. M. Gagné, B. S. Bloom, D. P. Ausubel, L. S. Briggs, G. A. Salomon, L. J. Cronbach, L. L. Thurstone, R. M. W. Travers, R. E. Clark, B. F. Skinner, C. E. Osgood, and G. L. Gropper are some of the researchers who have followed this line of research.

A second area deals with observational and/or sociological research. Researchers in this area acknowledge the notion that individuals can learn to perform some physical and social tasks by imitating the overt behaviors of a "model." This research area includes the humanistic, cultural, ethic, ethnic, ego and consistency needs of individuals when they are alone or in groups. This area also encompasses instructional cognitive styles which may be cultural and social preference systems acquired and supplemented during schooling (Heidt, 1977). Some of the proponents of the sociological approach to instruction include the following: A. Bandura, G. F. Kuder, E. K. Strong, G.

W. Allport, H. A. Witkin, N. Flanders, W. F. Seibert, R. E. Snow, R. R. Sears, A. H. Maslow, C. Buhler, and E. H. Erikson.

The third area is defined as physiological research. Within this area lies the subjective research on perception, on the form and structure of sensory messages, and on the constant interaction between the person and the environment. The biological bases of knowledge contain the roots of this major area. Research in this domain has dealt with the developmental characteristics of individuals as they interact, through the sensory channels, with the instructional environment. S. H. Bartley, C. B. DeSoto, J. J. Gibson, M. L. Fleming, J. Piaget, A. Gesell, R. J. Havighurst, H. Werner, M. Montessori, D. Durkin, V. Lowenfeld, F. M. Dwyer, and A. A. Lumsdaine are some of the researchers who have contributed to this area of study.

It appears that an appropriate research model must be directly related to the psychological, sociological, and physiological attributes of the learner, the teacher, the task, and the resources, which collectively result in an instructional environment. A Gestalt solution would be a research model, applicable to field studies in educational institutions, which accounts for the total instructional setting. If learning is the ultimate product of the instructional environment, then the instructional environment is the product of the interaction within, between, and among the teacher, the learner, the task, and the resources. One of the problems involved in a discussion of our instructional model is that the variables are not, unfortunately, as mutually exclusive as we would like for them to be. We are viewing instruction as a dynamic process in which the variables of instruction play an integral but subordinate part. Of greatest importance are the unique psychological, sociological, and physiological relationships within, between, and among the variables. We maintain that relationships between stimuli and responses are best predicted from information about the intermediary processes that occur within the individual. It is not unreasonable to suppose that the learner has developed general dispositions for processing stimulation based upon the daily activities associated with communicating, perceiving temporal and spatial relations, and problem solving. The learner must adapt himself to the learning environment in order, ultimately, to learn. Our model, then, is deduced from the psychological, sociological, and physiological makeup of the learner and his surroundings. To use Dale's (1969) "Cone of Experiences" as a simplistic example, if the learning environment is too "concrete," the learner will be under stimulated; if overly "abstract," the learner will be overwhelmed. In either case, it is quite likely that the learner will not reach the objective of the learning task. The implication sought here is that, in order to provide a functional relationship within, between, and among the variables of instruction, communications problems involving syntactics (interrelations of signs), semantics (meanings attributed to signs), and pragmatics (human reactions to signs) must be minimal. Hence, the objective form (physiological) and subjective meaning (sociological) of the learning task must yield a functional distinctiveness (psychological) in terms of the sensory information to be extracted by the learner.

Internal consistency within each variable of instruction is achieved only when the psychological, sociological, and physiological attributes are encoded and decoded in harmony. While it may be unrealistic to attempt to reduce the complexities of human nature to purely numerical terms, it seems worth emphasizing that, if viewed in the manner we have described, there is a potential of seven interactions within each variable of instruction for any given learning task. Carried further, the potential interactions among the variables of an instructional environment, comprised of the learner, the teacher, the task, and the resources, may be derived, in conservative mathematical terms, as approximately three million total interactions (Hoel, Port and Stone, 1971). Since these permutations are based upon one learner, rather than a class or cell of twenty or more learners, the implications are explicit. At any rate, these concomitant considerations serve to illustrate the complexity of any given instructional environment. The complexity is magnified still further by the fact that perceived dynamic internal and external attributes, that seem to be uniquely associated with the variables of the learning environment, are not always accurate reflections of the actual unique attributes. Over time, and through controlled research, the actual unique attributes for different learning environments may mainfest themselves as a subset of the perceived attributes. There is still, however, an additional complication, i.e., that either perceived or actual attributes may evolve, change, or disappear during the course of instruction or experimentation either due to maturation or due to interaction with other elements of the instructional environment. In the light of these considerations, it seems particularly apt to note that Cronbach (1975) said ". . . the line of investigation I advocated in 1957 no longer seems sufficient. Interactions are not confined to the first order; the dimensions of the situation and of the person enter into complex interactions" (p. 116).

Where Do We Go From Here?

It would seem that the need exists for an instructional model that incorporates the past, present, and future researchable directions. This is not to say that future researchers will agree that the postulated relationships in vogue today are researchable tomorrow. Nonetheless, it may be possible to use the model discussed hereinbefore as a foundation or datum plane upon which a logical rationale could be based for future meta-analyses. At the same time, psychological, sociological, and physiological measuring instruments and/or inventories with veridical comparisons should be identified and/or developed in an attempt to differentiate the unique attributes from the common and the static attributes from the dynamic. As a result, functional experimental research investigating the interrelationships of the variables of instruction will, in the future, be equipped to employ realistic rational controls so that experimenters may more reliably explain that which actually happened.

Most contemporary researchers would agree that we need to know more about the physical and psychological attributes of resources. Heidt (1977) suggested that the unique psychological attributes of resources may be a product of the physical attributes for specific learning experiences. We submit that a similar subset of unique sociological attributes should also be specified since they influence, and are influenced by, the unique psychological and physiological attributes. To extend this idea, the subsets should also be comparable to all of the variables of instruction, and not just to the resources.

In conclusion, it seems reasonable to believe that a logically deduced amalgamation of all research in the behavioral sciences could result in one or more axiomatic theories for instruction. Once derived, selected constituent attributes could be held constant while, at the same time, systematically varying others. Until this is accomplished, research on instructional media will remain omnibus, composed of complex and multivariate aspects of what might be termed "impulsive reckoning."

Selected References

Allen, W. H. Instructional media research: Past, present and future. *AV Communication Review*, 1971, *19*, 9-18.

Allen, W. H. Intellectual abilities and instructional media design. *AV Communication Review*, 1975, *23*, 139-170.

Anderson, S. B., Ball, B. and Murphy, R. T. *Encyclopedia of Educational Evaluation.* Washington, D.C.:Jossey-Bass Publishers, 1975.

Becker, A. D. Alternate methodologies for instructional media research. *AV Communication Review*, 1977, *25*, 181-194.

Berliner, D. C., and Cahen, L. S. Trait-treatment interaction in learning. In F. N. Kerlinger (Ed.), *Review of Research in Education: 1.* Itasca, Ill.:F. E. Peacock, 1973.

Bracht, G. H. Experimental factors related to aptitude-treatment interactions. *Review of Educational Research*, 1970, *40*, 627-645.

Burns, R. W. Instructional television, interaction, and learning objectives. *Educational Technology*, 1976, *16*, 44-49.

Carpenter, D. R. The application of less complex instructional technologies. In W. Schram (Ed.) *Quality in Instructional Television.* Honolulu:The University Press of Hawaii, 1972.

Clark, R. E. Constructing a taxonomy of media attributes for research purposes. *AV Communication Review*, 1975, *23*, 197-215.

Comstock, G. *Television and Human Behavior: The Research Horizon, Future and Present.* Rand Corporation Publication R-1748-CF (Santa Monica, California, June 1975).

Conway, J. K. Forms of information representation and cognitive operations. In G. Salomon and R. E. Snow (Eds.) *Viewpoints* (Bulletin of the School of Education, Indiana University, 1970, *46*, 154-164.

Cronbach, L. J. The two disciplines of scientific psychology. *The American Psychologist*, 1957, *12*, 671-684.

Cronbach, L. J. Beyond the two disciplines of scientific psychology. *American Psychologist*, 1975, *30*, 116-127.

Cronbach, L. J. and Snow, R. E. Individual differences in learning ability as a function of instructional variables. *USOE Final Report*, Contract No. OEC-4-6-061269-1217. Stanford:Stanford University, School of Education, 1969.

Cronbach, L. J. and Snow, R. E. *Aptitudes and Instructional Methods.* New York:Irvington, 1977.

Dale, E. *Audiovisual methods in teaching.* Third edition. New York: Holt, Rinehart and Winston, 1969.

DiVesta, F. J. Trait-treatment interactions, cognitive processes, and research on communication media. *AV Communication Review*, 1975, *23*, 185-196.

Dwyer, F. M. *A Guide for Improving Visualized Instruction.* State College, Pa.:Learning Services, 1972.

Dwyer, F. M. *Strategies for Improving Visual Learning.* State College, Pa.:Learning Services, 1978.

Fleming, M. L. Media research prospects: Considerations from a perceptual point of view. In G. Salomon and R. E. Snow (Eds.) *Viewpoints* (Bulletin of the School of Education, Indiana University), 1970, *46*, 119-136.

Glaser, R. Indivduals and learning: The new aptitudes. *Educational Researcher*, 1972, *1*, 5-13.

Hawkridge, D. G. Media taxonomies and media selection. Paper prepared for the Institute of Educational Technology. The Open University, Milton Keynes, England, 1973.

Heidt, E. U. Media and learner operations: The problem of a media taxonomy revisted. *British Journal of Educational Technology*, 1977, *8*, 11-26.

Hoel, P. G., Port, S. C., and Stone, C. J. *Introduction to Probability Theory.* Boston:Houghton Mifflin Company, 1971.

Levie, W. H. and Dickie, K. E. The analysis and application of media. In R. M. W. Travers (Ed.) *Second Handbook of Research on Teaching.* Chicago:Rand McNally College Publishing Company, 1973, 858-882.

Mielke, K. W. Research and evaluation in educational television. In A. Coldarci and R. Arnove (Eds.), *Educational television: A policy critique and guide for developing countries.* A report to the Ford Foundation, Stanford, California:School of Education, Stanford University, 1973.

Ott, L. *An Introduction to Statistical Methods and Data Analysis.* Belmont, California:Wadsworth Publishing Company, Inc., 1977.

Parkhurst, P. E. Generating meaningful hypotheses with aptitude-treatment interactions. *AV Communication Review*, 1975, *23*, 171-196.

Salomon, G. What does it do to Johnny? In G. Salomon and R. E. Snow (Eds.) *Viewpoints* (Bulletin of the School of Education, Indiana University), 1970, *46*, 33-62.

Salomon, G. Heuristic models for the generation of aptitude-treatment-interaction hypotheses. Paper presented at the Meeting of the American Educational Research Association, New York, February 1971.

Salomon, G. A cognitive approach to media. *Educational Technology*, 1976, *16(5)*, 25-28.

Salomon, G. and Clark, R. E. Reexamining the methodology of research on media and technology in education. *Review of Educational Research*, 1977, *47*, 99-120.

Salomon, G. and Snow, R. E. Summary and some reflections on the conference. In G. Salomon and R. E. Snow (Eds.) *Viewpoints* (Bulletin of the School of Education, Indiana University), 1970, *46(5)*, 183-199.

Schramm, W. *Big Media, Little Media.* Beverly Hills, Ca.:Sage Publications, Inc., 1977.

Shulman, L. S. Reconstruction of educational research. *Review of Educational Research*, 1970, *40*, 371-396.

Snow, R. E. Research on media and aptitudes. In G. Salomon and R. E. Snow (Eds.) *Viewpoints* (Bulletin of the School of Education, Indiana University), 1970, *46(5)*, 63-89.

Snow, R. E. and Salomon, G. Aptitudes and instructional media. *AV Communication Review*, 1968, *16*, 341-357.

Virag, W. F. Effectiveness of three instructional modes employed to transmit content to students with different aptitude patterns. In P. J. Sleeman and D. M. Rockwell (Eds.) *Instruction Media and Technology*. Stroudsberg, Pa.:Dowden, Hutchinson and Ross, Inc., 1976.

RESEARCH INTERPRETATION: A JUSTIFICATION FOR MEDIA

Dr. Francis M. Dwyer
Senior Research Associate of Education
114 Mitchell Instructional Services Bldg.
The Pennsylvania State University
University Park, Pennsylvania 16802

Presented Research & Theory Division
Association for Educational Communication Technology
New Orleans, Louisiana, March 5-9th., 1979

Research into Practice

(Presentation Abstract)

RESEARCH INTERPRETATION: A JUSTIFICATION FOR MEDIA

This paper will attempt to explain the apparent inconsistencies in media research findings by identifying apparent research weaknesses which complicate data interpretation and frustrate attempts to derive broad generalizations useful to practitioners in their classroom use of visual media. The presentation will emphasize the interrelationships which exists among the different types of visual materials, different types of educational objectives, different methods of presentation, different cueing techniques, etc. Studies describing the results obtained from evaluations involving television, slide/audiotape, programmed instruction, and textbook formats will be reviewed.

RESEARCH INTERPRETATION: A JUSTIFICATION FOR MEDIA

Education is currently experiencing financial crises which are necessitating that virtually everyone justify the financial outlay made for his position, equipment, and materials. At all levels instructional media specialists and users of educational technology are being challenged to defend their usefulness in the teaching-learning process. This is not always easy to do because of the contradictory nature of empirical findings which have resulted from media related research studies. For example, if one were to survey the available experimental media research, a number of studies would be found in which instruction utilizing media was significantly more effective than conventional instruction. Some studies would also be found in which the mediated instruction was found to be significantly less effective than conventional instruction, and others would be found in which no differences were found to exist between the mediated and conventional instruction. To further complicate media justification, when significant differences are obtained in empirical studies, the results are seldom in agreement with other research findings investigating similar problems (Davis, 1962; Wendt & Butts, 1962; Stickell, 1963; Reid & MacLennan, 1967; Chu & Schramm, 1967; Edling, 1968). Opponents of media and technology are quick to cite these contradictory findings as an indication of ineffectiveness.

Since it has been documented that properly designed visual materials can significantly improve student achievement (VanderMeer, 1950b; Kopstein & Roshal, 1954; Treichler, 1957; Gropper, 1962; Dwyer, 1972, 1978), it is imperative that individuals charged with the responsibility of justifying media utilization be able to detect those deficiencies in media research which contribute to its failure to generate consistent findings of superiority. After deficiencies are identified, contradictory findings can be explained and the use of media justified.

Following is a list of some of the most common criticisms associated with media research which tend to complicate data interpretation and frustrate any attempts to derive broad generalizations useful to practitioners in the classroom (Dwyer, 1978 pp. 39-41):

1. Many studies have obvious weaknesses in experimental design, i.e., lack of randomization of students, vagaries in sampling, inadequate numbers of students in treatment groups, lack of tests of significance and probability statements, absence of any control factors.
2. A considerable number of media related studies which have been reported are without any hypotheses or predictions based on theory.
3. The content material being presented experimentally in media related studies has been restricted in that it is far removed from that which is currently being taught in schools (nonsense syllables, digits, letters, etc.); consequently, the results have had little practical significance to educators involved in applications in the classroom.
4. The difficulty and meaningfulness of the content material (in terms of the kinds of educational tasks to be achieved by the students) used in the experimental treatments has not been specified precisely.
5. In many media related studies content to be taught has not been pretested to determine where visualization of the the content is appropriate; in other words, a considerable amount of visualization used in media related experiments has not been specifically designed to complement the content material to be presented.

6. In media related research the relationship that exists between the content information in each channel (visual, oral, print) to the information in the other channel(s) needs to be specified precisely.

7. The method of presenting the visualization has not been described precisely.

8. The precise purposes of visualization in many mediated studies has not been stated.

9. The amount of realistic detail (line drawings, detailed, line drawings, photographs) contained in visualization designed to complement oral/print instruction in media research has not been described precisely so that data from different studies can be compared.

10. Mediated instruction normally is evaluated via printed criterion items. To properly assess the instructional contribution of visualization, a significant proportion of the evaluation battery should reflect all the channels (visual, print, oral, etc.) used by the learner to acquire the information.

11. Many of the criterion tests currently being used to measure achievement are global in nature attempting to measure the students' total learning, rather than being designed to measure media's contribution to students' achievement of specific kinds of educational objectives. In using the global criterion test to measure student achievement of different types of objectives, the variances are pooled (for the different objectives), thereby concealing any effects that media may have in faciliating student achievement of specific educational objectives.

12. Many of the criterion measures used to assess student information acquisition may be invalid or unreliable--reliability and validity coefficients are very infrequently reported in media related research.
13. The type of test items used to measure student information acquisition has not been adequately specified. This is important because item format can influence the student's level of achievement; i.e., a fill-in type test in which the student is required to provide the correct word is more difficult than a recognition type multiple-choice type test in which the student is required to select the correct response from an array of possible alternatives.
14. Researchers have failed to specify the time-span (seconds, minutes, hours, etc.) between the presentation of the information to the students and the testing.
15. The amount of time students have been permitted to view and interact with the instructional presentations has not been adequately controlled and/or reported.
16. Media researchers in preparing their research for publication have omitted essential characteristics of their studies which prevents exact replication, i.e., total time of treatment, age or grade level of students, mean intelligence of students, prior knowledge in the content area, etc.
17. Very few of the media related studies have been replicated to establish confidence in the results.
18. The diversity of interests and a lack of common usable terminology used by media researchers has compounded the misinterpretation of many media related studies and restricted their generalizability.

19. To a certain extent the results of experiments have been determined by the statistical techniques used--a liberal test providing one interpretation of the results, a conservative statistic another.

REFERENCES

Chu, G. C. & Schramm, W. 1967. Learning from television: what the research says. Washington: National Association of Educational Broadcasters.

Davis, O. L., Jr. 1962. Textbooks and other printed materials. Rev. Educ. Res. 22: 127-140.

Dwyer, F. M. 1972. A guide for improving visualized instruction. State College, Pennsylvania: Learning Services, Box 784.

Dwyer, F. M. 1978. Strategies for improving visual learning. State College Pennsylvania: Learning Services, Box 784.

Edling, J. V. 1968. Educational objectives and educational media. Rev. Educ. Res. 38: 177-194.

Gropper, G. L. 1962. Studies in televised instruction: the role of visuals in verbal learning. Pittsburgh: Metropolitan Pittsburgh Educational Television Station and the American Institute for Research.

Kopstein, F. F. & Roshal, S. M. 1954. Learning foreign vocabulary from pictures vs. words. American Psychologist. 9: 407-408.

Reid, J. C. & MacLennan, D. W. 1967. Research in instructional television and film. Washington, D.C.: U.S. Office of Education. HEW Bureau of Research.

Stickell, D.W. 1963. A critical review of the methodology and results of research comparing televised and face-to-face instruction. Doctoral dissertation. The Pennsylvania State University.

Treichler, D.G. 1967. Are you missing the boat in training aids? Film & AV Commun. 1: 14-16.

VanderMeer, A.W. 1950. Relative contributions to factual learning of the pictorial and verbal elements of a filmstrip. Sch. Rev. 58: 84-89.

Wendt, P. R. & Butts, G. K. 1962. Audiovisual materials. Rev. Educ. Res. 32: 141-155.

Gary L. Filan
Educational Technology
Box FLS
Arizona State University
Tempe, AZ 85281

Behavioral Objectives: A Critical Review of Theory and Research

Gary L. Filan

Vernon S. Gerlach

Abstract

Behavioral objectives provide the basis for systematic planning of instruction. This systematic approach enables the designer to work more effectively and it enables the learner to understand what is expected upon completion of the learning experience. There are, however, a wide range of views concerning the advantages of behavioral objectives. The purpose of this paper is to present a critical review of the literature on behavioral objectives. This paper is divided into four major areas: (1) definition of a behavioral objective; (2) the function of behavioral objectives; (3) a consistent acceptable format for constructing behavioral objectives; and (4) the pros and cons of behavioral objectives.

The literature review reveals that current findings on the effects of instructional objectives provide no conclusive or consistent data on the relationship between the use of objectives and student learning. Consequently, there is a need to assess the behavioral objectives movement, to identify strengths and weaknesses, and to suggest areas in which research is needed.

Introduction

Behavioral objectives have been central to the concept of instructional systems development. They have now been incorporated into the designing of curriculum. They provide the basis for planning instruction. They have been used to tell learners what is expected of them upon completion of the learning experience. There is, however, a wide range of views concerning the advantages of behavioral objectives, as well as many varying opinions as to the technical aspects of how and for what purpose they should be used. Behavioral objectives provide a point of departure for a thoroughgoing attempt to improve instruction. By precisely stating in behavioral terms what the student should be able to do after the learning experience, the designer hopes to reduce any gaps between the desired outcomes of education and the intentions of the instructor. This approach has been criticized from both curriculum specialists and educational technologists. This paper is a critical review of the literature on behavioral objectives; it is divided into four parts: (1) determining a consistent definition, (2) a consistent acceptable format used for writing behavioral objectives, (3) the function of behavioral objectives, and (4) the cases for and against the use of behavioral objectives.

I Defining Behavioral Objectives

At first it may seem that defining objectives is really not a difficult task. However, educators experienced with curriculum development, course development, and those who have tried to develop procedures for evaluating students will attest to the fact that it is. Palmer (1974) feels that most educators make the task of defining behavioral objectives far too complicated.

This section of the paper will be divided into five parts: (1) Defining Behavioral Objectives in Relation to Terminal Behavior, (2) Defining Behavioral Objectives in Relation to Subject Matter, (3) Operationalism and Behavioral Objectives, (4) Opposing View Points Concerning the Definition, and (5) Research Related to the Development of a Consistent Operational Definition.

Defining Behavioral Objectives in Relation to Terminal Behavior

Many educators have defined the term behavioral objective. Lindvall (1964) states that the process of developing behavioral objectives is basically one of facilitating communication. This is accomplished by choosing precise words and statements so that there is a clear and exact meaning for those reading the objective. Popham (1969) writes that whether these statements are referred to as objectives, aims, goals, intents, or outcomes is relatively unimportant. Whatever synonym is used, a behavioral objective should refer to an intended change which one wishes to bring about in a learner. Bloom (1956) defines objectives as being specific formulas that the educative process uses to change student behavior. Mager's (1962) definition of behavioral objective has probably influenced more educators than any other definition: an objective is a statement describing a proposed change in a learner; it specifies what the learner will be like when he has successfully completed a learning experience. For example, an objective written for a 5th grade science class using Mager's 3 characteristics of a well-stated objective would look like the following: Given a battery, light bulb, socket, and pieces of wire, the student will be able to demonstrate the making of an electric circuit by connecting wires to battery and socket and testing the lighting of the bulb.

Defining Behavioral Objectives in Relation to Subject Matter

A major consideration when one is defining behavioral objectives is that of determining what is to be learned by the learner. An educational objective has been described as one in which the learner's behavior is clearly and precisely specified in relation to the subject matter with which the learner is expected to deal. That is, the objective must specify not only the learner's terminal behavior, but also the particular aspects of the subject matter to which the learner must address himself in order that learning may occur. Gagné and Briggs (1974) state that the first step in defining objectives is to identify the purpose of the course. This purpose should be concerned with what behavioral change will take place should the purpose of the course be attained. They also feel that these purposes should be stated as immediate outcomes of instruction, and not outcomes to be reached in the distant future. This process of identifying the purpose will help teachers to make clear statements of what they are trying to teach. Some teachers in the past have had a clear understanding of what was to be taught and what was to be learned by the student, and were able to translate this notion into relevant learning experiences without ever having put them down on paper. However, many other teachers have not carried their thinking beyond the point of selecting the content to be presented. They have not considered carefully what the students are to do with the information.

Operationalism and Behavioral Objectives

Operationalism is a concept borrowed from the hard sciences. It is concerned with banishing ambiguity and obscurity from the language of science. By applying scientific concepts to concrete procedures one could avoid inconsistent and contradictory meanings. In the context of objectives it refers to the process of defining abstract constructs or concepts in terms of a limited number of instances drawn from the three domains of learning: cognitive, affective, and psychomotor (Tieman, 1977). Tuckman (1972) says an operational definition is a definition based on the observable characteristics of that which is being defined. In the field of behavioral research, operational definitions are formulated so that statistical methods can be applied. These methods produce reportable evidence and hard conclusions. The behavioral objectives approach requires that behavioral objectives be precisely stated in behavioral terms.

Opposing View Points Concerning the Definition

Some educators feel that there is difficulty with explicitly defining behavioral objectives. MacDonald-Ross (1973) points out that some of the problems encountered in the behavioral objective domain are extensions of the basic problems faced by operationalism. He states: "What exactly counts as an operation? What happens to the concepts when we are not performing operations or if we have not yet learnt how to perform them?" Hempel (1958) says that the greatest advances in scientific systematisation have not been accomplished as a result of referring explicitly to observable behaviors, but rather by means of laws that speak of various hypothetical or theoretical attributes. He points out that activities, events, and attitudes which are not ascertainable by direct observation have an important and valid place in the educational system. For instance, in the fine arts it is extremely difficult to have an observable product when judgment, feeling, and creativity play such a major role. MacDonald-Ross points out that as far as art subjects are concerned, there are no ultimate goals to be reached, but rather standards of judgment and tastes to be developed. He also says that these broad goals in the arts do represent a type of behavior, which being internal is not observable. Eisner (1967) supports the positions that attitudes, values, and creative experiences are important educational aims which cannot be translated into behavioral terms. Burns (1972) feels that if the definition of behavioral objectives is concerned only with specific behaviors, there is no room for expansion, self discovery, originality, and whatever you might wish to call that which is subsumed under the general term "creativity."

Research Related to the Development of a Consistent Operational Definition

A series of studies (Barron, Gerlach, and Haygood, 1976, and Haygood, Gerlach and Wigand, 1977) deal with analyzing rater's perception of the components of behavioral objectives, rated both in isolation and within complete statements of objectives. These studies measured the degree to which the various components contribute to the raters' perceptions of the complete objective. These empirical studies have currently investigated the development of a consistent operational definition of the terms behavioral objective. The results indicate clearly that no single component, the verb, direct object, condition, or standard, should be singled out as being of primary importance in determining the character of a behavioral objective. Investigators are moving closer towards a consistent operational definition of the behavioral objective, but additional research is needed that will limit the many discrepancies among educators concerning the definition.

II Form

Major Theorists Views on Form

Many articles and books have appeared in the professional literature concerning the proper form of behavioral objectives (Mager, 1962, Bloom, 1964, Lindvall, 1963, Popham and Baker, 1970, and Kibler, Barker, and Miles, 1970). Tyler (1934) suggests one should state the objectives in such clear and definite terms that they can serve as guide for constructing test questions. Many statements of objectives are so vague and nebulous that they prove to be glittering generalities which are of little value as guide in teaching and of no value in making examinations. Mager's (1962) three criteria for a well stated behavioral objective are probably the best known: (1) One should state the objective in terms of what the learner will be able to do after the learning experience. This is done by selecting verbs which describe observable actions. Such words as identify, describe, construct, and list are far less ambiguous than verbs such as to know, understand, or appreciate. (2) The second characteristic of a well stated objective is a statement of the conditions under which the performance is to occur. Conditions should be stated clearly enough that others understand your intent as you understand it. (3) The third characteristic of a well stated objective is the criterion, the quality or level of performance that will be considered acceptable.

Some educators feel that Mager's criteria for a well-stated behavioral objective have weaknesses. Merrill (1970) reports that Mager's criteria for a well-stated objective fail to distinguish the level of behavior. He states that there are more purposes to instructional objectives than transmission of knowledge and increasing proficiency. He also points out that there are two classes of conditions under which behavior is to occur. The first is concerned with those conditions related to a particular subject matter and unique to the testing situation. An example of a condition stated in a behavioral objective for a math class would be "...using only a calculator..." or "...using only the protractor..." The second is concerned with the psychological conditions which help define the behavior being observed. This second type is quite often overlooked and is more important because the type of behavior being observed will change when psychological conditions are changed. In most cases the psychological conditions are not stated in the objective, but have an important effect upon its outcome. For example, the classroom learning environment is typically not normal the day before Christmas vacation. MacDonald-Ross (1973) feels that a fourth characteristic should also be considered when determining what constitutes a well stated objective: an objective should be relevant to the general educational aims of a course. He states, "No rules are given for achieving this criterion--which is actually the most difficult to achieve, yet the most important of all" (p. 12).

Gagné and Briggs (1974) agree with the three basic criteria set forth by Mager and later writers concerning a well-stated behavioral objective. Gagné and Briggs also state that the choice of verb in an objective is a matter of critical importance. They feel that there are two kinds of verbs which must be incorporated into an objective. The first verb denotes action. Verbs denoting action are not difficult to find. Common ones are, writes, draws, selects, matches, names, groups, verifies. There are many others as well. The following examples denotes action: Without use of reference materials, state the provisions of the Fifth Amendment, in writing. While it may be essential for completeness of communication, is not necessarily the most important verb in the definition of an objective. The second verb, (i.e., the major verb) which they feel is probably of even greater importance in its implications, denotes learned capability. It has the purpose of communicating the kind of human capability one expects to be learned, as it may be observed in some performance exhibited by the student. The following verbs describe performances implying learned capabilities: discriminate, classify, demonstrate, generate, execute, originate, identify, and state. Several examples that use verbs which describe learned capabilities are: "...states orally the major issues in the Presidential campaign of 1968," and "...identifies, by naming, the root, leaf, and stem of representative plants." Early writers regarded the verb as the primary determiner for which objectives were considered behavioral. Many writers provided us with lists of verbs. Not until Deno and Jenkins (1968) was there any empirical data collected regarding the verb and behaviorality. Deno and Jenkins selected a list of verbs from a well-known experimental curriculum. They had a group of elementary and secondary teachers rate the verbs on a five point scale of observability. The following results were reported by Deno and Jenkins "The results indicate that many widely used and recommended behavioral terms refer to behavior which is not regarded by teachers to be as clearly 'observable' as some have suggested" (p. 22). They concluded that verbs used in behavioral objectives are selected for usage rather than observability. Gerlach (1974) replicated the Deno and Jenkins study, by rating the same ninety-nine verbs. The results obtained by the study were basically the same as those reported by Deno and Jenkins.

A Closer Look at the Three Essential Characteristics of a Behavioral Objective

A verb which describes overt behavior is the main factor involved in stating clear descriptions of what the learner must do to perform the task. There are many verbs which could be used in behavioral objective statements. The following list is illustrative of widely used verbs: identify, name, describe, construct, state, discriminate, classify, generate, name, order, check, and perform (Sullivan, 1969, Deno and Jenkins, 1968, Gagné and Briggs, 1974, Gerlach, 1974).

The choice of verb in an objective is a matter of critical importance. The primary reason is the avoidance of ambiguity. The statement of an objective should communicate reliably, in such a way that two different literate people will agree on the specific behavior which is to be exhibited by the learner. Words such as "knows," "understands," "appreciates," do not communicate reliably. The action should be expressed in the objective so that anyone who reads it will be able to identify the same performance.

The statement of conditions which specifies the condition under which the behavior is to occur is the second essential characteristic of a well-stated behavioral objective. The conditions specify the limitations and restrictions which are imposed on a learner when, following instruction, he performs the task stated in the objective. Conditions describe the materials, events, information, and the objects in the learner's environment. Examples of stimulus conditions include the following:

"When presented with a typed list..."

"With the use of class notes..."

"Without the use of classnotes or other references..."

Ambiguity is reduced when precise limitations and restrictions are specified.

The third essential characteristic of a well stated behavioral objective is the statement of criterion, which describes how well the learner is to perform the task. The criterion or standard provides a basis for evaluating the prescribed behavior. For example, consider the objective "Name the four major food crops grown in Arizona." The standard is "correctly name all four major food crops grown in Arizona and only those four. Thus a performance standard is a specified level of achievement used to determine whether or not a task has been mastered satisfactorily. Performance standards help both teachers and students know where any given student is in a program. Mager (1962) states, "When the minimum acceptable performance for each objective is specified we have a performance standard to use in assessing students' work.

Mastery generally means that the student will exhibit the performance 100% of the time (minus some small percent for "measurement error"). However, frequently it is appropriate to set a lower standard, such as three out of five problems solved correctly, or four out of six defects identified (Bloom, 1971). Briggs (1970) states, "Many people find the how well criterion the most awkward to include in a statement of objectives. But, for objectives requiring more complex evaluation, it may be easier to omit this third criterion from the actual statement of the objective, and present it in the scoring key, and grade conversion guide, showing just what standard of student performance will be considered acceptable."

Behavioral objectives do not state in quantitative terms what criteria will be used to determine whether or not the objective has been satisfactorily met (Gagné and Briggs, 1974). The objective does not say how many times the student is to "demonstrate the addition of whole numbers," or how many "errors" will be permitted. They do not state what will be needed for the observer to be confident that the objective has been met. Gagné and Briggs feel that there are two reasons why the criteria should not be included in the objective statement. First, the criteria specified in an objective is not likely to be applied in the same manner to all individuals. Second, the question of criteria of performance is a question of "how to measure," and is bound up with the techniques of performance assessment. At the point in the instructional planning when objectives are being described, it is confusing to become concerned with assessment procedure. The concept of mastery implied by the objective statement is derived from an important theoretical viewpoint. The theory underlying Gagné's (1970) learning hierarchies accounts for the function of mastery. According to the theory, the achievement of an intellectual skill is important because it supports the learning of more complex skills. Strictly speaking from a practical view, Gagné and Briggs point out that it is not possible to predict in precise terms how mastery should be measured. They state, "It is not wise to adopt some arbitrary standard like five out of six correct responses. The criterion of mastery will vary with what is being learned, and needs to be determined as a part of the assessment process" (p. 89).

The roles of the three basic components of an objective, as stated by Mager (1962) have been researched by Barron and Gerlach (1974). Their results confirmed the importance of the verb in objectives, but they also found that the choice of conditions and criteria influences the rating of a complete objective. Haygood et al. (1977) state that "no single component, such as the verb, should be singled out as being of primary importance in determining the characteristic of a behavioral objective."

Some writers contend that form should be considered only as a function of an objective. There is little point in requiring a teacher to write an objective in standard form without taking into account the purpose for stating the objective (Harlen, 1972). This purpose should provide the basis for teachers to make decisions in their everyday work in guiding learning in the classroom. Of course there will be a variation from one teacher to another as to the form of expression. Harlen states, "The form in which the objectives are stated must also be left to the teachers; trying to specify them in the detail advocated by Mager may be of help to some in encouraging clarity of thought, it may be unnecessary for others" (p. 234).

Instructional, or classroom, objectives are primarily the responsibility of teachers and cannot be determined by anyone else (Harlen, 1972). Unfortunately, teachers are frequently untrained in the use or formulation of behavioral objectives in which case outsiders suggest examples or provide guides to defining objectives. This may do as much harm as good if teachers accept others' objectives as their own, or if they go through a superficial training which teaches them the form but not the philosophy behind the concept of objectives (Marlen, 1972).

Many teachers are now being given the opportunity to implement a developed objective based instructional program. Niedermeyer and Sullivan (1977) state that teachers do have the option of accepting or rejecting an objective based program on their judgment of the worth of its objectives and resources for their pupils. There is no explicit requirement that teachers are to use all the materials that have been developed for them, or closely follow the recommended instructional procedures. The intention of the teacher and the program, however, is to produce successful pupil performance on the objectives. To reach this goal teachers should use whatever resources and creative abilities they have.

III Function of Behavioral Objectives

The functions of behavioral objectives can be divided into four categories: (1) aid in design of developing efficient instructional programs; (2) provide guidance in evaluation of instructional programs; (3) facilitate learning for students; and (4) inform teachers, administrators, and general public of the purposes of the instructional program.

Aid in Design of Instruction. Objectives offer a systematic means of planning in education. When designing a program or system, one needs to know what a successful solution will look like as well as what criteria it must satisfy. MacDonald-Ross (1973) implies that behavioral objectives can provide the only possible rational basis for evaluating the success of the learning experience. The course becomes successful only if the students can demonstrate satisfactorily what the objectives predict. He also states that objectives indicate how the process of teaching should be conducted as well as help to assist in the selection and design of instructional activities.

A systematic procedure for developing instruction has been developed by Gagné (1974). He states that when objectives are known, one is able to infer what kind of learned capability is being acquired, and one can also determine what conditions will be needed to bring about the learning with greatest efficiency. Clearly, then, the systematic design of lessons which make up courses will result in the development of a sizeable collection of statements of objectives. This collection of objectives will be constructed by using such schemata as Bloom's Taxonomy (1956) or Gagné's learning hierarchy (1970). Higher level objectives will be formulated which will depend on the acquisition of lower level objectives. These lower level objectives will be stepping stones or prerequisite skills that will have to be mastered before the higher level goal or objective can be achieved. Thus the specification of prerequisite skills should provide a complete description of those previously learned skills needed by the learner in order to acquire the new skill most readily. The identification of performance objectives makes possible the classification of capabilities into useful categories. Without these categories, we can deal with learning principles only on a very general basis. With them, it becomes possible to infer what kinds of learned capabilities are being acquired at any given point in the learning process. One can also determine under what conditions (internal as well as external) the learning experience takes place. This knowledge may increase the efficiency of one's instruction.

Instruction is to be designed with reference to performance objectives and the prerequisite capabilities they imply. The first necessary component in the design of instruction is to classify the lesson as having a particular type of learning objective. Along with specifying objectives, two other components are included in the design of instruction. One is developing methods, designing materials; producing media, and developing learning experiences or exercises; the other is evaluating the success of the learners after the instructional process. Mager (1968) describes the three components of instruction in an easy to remember format: (1) Where am I going?, which refers to how to achieve the objectives; (2) How will I get there?, which refers to how to achieve the objectives; and (3) How will I know when I've arrived?, which is the evaluation process of determining whether or not the student has satisfactorily achieved the objective. These three questions can be used when planning instruction.

One does not have to proceed in any given order when developing the three components. MacDonald-Ross (1973) feels this would be entirely too mechanical a view of the procedure of instructional design. Rather, he suggests that the designer should do his best in developing objectives, then move on to considering the end-of-unit tests, and then select and develop the instructional materials. This procedure would be carried out until one felt that each component had been specified as clearly as possible. At this point one would develop a first draft; quite frequently there will be changes in some of the objectives. But of course one would be doing this on the basis of some evidence rather than on the basis of some vaguely conceived or haphazard scheme. One should realize that although objectives are intended to be a basis for prescribing course structure and evaluation, frequent adjustments must be made in practice.

There are varied viewpoints concerning the advantages and disadvantages of using behavioral objectives in instructional design. Baker (1974) for example, feels that the use of behavioral objectives in forming a basis for restructuring instructional programs may have some negative consequences. Because objectives are stated in operational language, they appear to be more teachable. Objectives may look achievable if they follow the formula: "Given...the student will be able to...", but such is not always the case. Because it is easy to transform goals into the accepted behavioral objectives format, examples of learning may be casually produced. Baker states that many supervisors and curriculum specialists feel that as long as the behavioral verb has been supplied, there is little to criticize. She also states that "most behavioral objectives do not present sufficient cues regarding what a teacher should alter in instruction in order to facilitate improved learning."

Objectives help as a stimulus to clear thinking by forcing the teachers to think in specific terms rather than in vague ambiguities. MacDonald-Ross (1973) feels that this is a prerequisite for any system of design or planning and that such thinking yields the additional benefit of revealing value judgments that might otherwise remain concealed. Once externalized, such thinking can be subjected to criticism and testing, and thus instruction can be improved. Since objectives can provide a stimulus for clear thinking, they can help teachers in developing instructional goals, strategies, purposes, and methods. Kibler, Cegala, Parker, and Miles (1974) suggest that if teachers state their instructional intent in behavioral objectives, other teachers will be able to understand what content is being covered within their classroom.

Although there is not complete acceptance among educators of the specific use of behavioral objectives in designing instruction, it can be agreed that behavioral objectives can provide guidelines for teaching and can lay the foundation for a systematic approach for curriculum planning.

Guidance in Evaluation of Instruction. Objectives are useful in the evaluation process. Gagné (1970) states that descriptions of objectives are descriptions of what must be observed in order to verify that the desired learning has taken place. Consequently, statements of objectives are used for assessing student learning. Teachers may use objectives to design situations within which student performance can be observed; or objectives can be used as a basis for test construction. While objectives can be used as a basis for evaluating students, they can also be used as a basis for evaluating instruction. Since objectives are directly related to instructional content, and since they include a performance standard, both the student and the teacher can know the quality and quantity of a successful performance. If students constantly fail to meet the standard specified in the objective, this can help the teacher to evaluate either the instructional content or activities that are related to the objectives not being attained by the students. Revision of the instructional content and/or activities may be needed at this time. Kibler et al. (1974) feel that there are primarily three functions of evaluation in instruction: (1) student achievement of instructional objectives; (2) evaluation of instructional materials; and (3) evaluation of the instructor. Kibler goes on to say that while both norm-referenced and criterion-referenced testing can be used to provide information concerning the three functions, criterion-referenced testing is best suited for accomplishing functions one and two, and norm-referenced testing is best suited for function three. It is extremely important that teachers be able to determine the student's level of achievement at any time during an instructional program. This can be done very effectively by keeping an accurate record of the students' progress on each performance objective throughout the instructional program. When teachers have this type of information, they will know how the student is performing at any time and they will be able to pinpoint any weaknesses in learners.

Evaluation can occur throughout the instructional process. Briggs (1970) states that tests over competencies of an objective are useful for determining whether or not students need additional remedial work. They also are a useful source for pinpointing trouble when a student fails the test of a specific behavioral objective. When tests are given for specific objectives, then can serve as a guide for the teacher in determining whether the student is ready to go on to the next objective. Tests for units of instruction can reveal the learner's mastery of more complex objectives. End-of-course tests can indicate the students' ability to solve more complex problems or to apply their knowledge to a wider range of situations. Gagné (1974) feels the pre-tests based upon criterion objectives can also help to identify students who have acquired the level of performance before instruction begins. Such objectives may also help in identifying students who lack the pre-requisites to satisfactorily meet the criteria set forth in an objective.

Teachers have few rules to go by when writing test items; selection of content is often haphazard. Consequently, when teachers are faced with a student who has not satisfactorily met the objectives, they often have difficulty in selecting content for practice items. Teachers usually guess, after inspecting the test, what relevant class of examples they may use that will correspond to the objective. Baker (1974) feels the answer to providing data to facilitate improvement of instructional programs lies in domain referenced testing. This type of testing can supply both the data needed for assessment of instructional programs and information suitable for feedback to teachers to facilitate planning. The use of domains in the design of tests helps reduce the production of trivial objectives. A domain consists of a subset of knowledge, skills, understanding or attitudes where the essential elements of the content, in which the student is expected to acquire, is carefully described. Baker states, "domains for teaching and testing represent an attempt to find a reasonable compromise between vagueness and over-precision" (p. 11). Domain require the teacher to focus on the range of eligible content to which the learner's skill is to apply. Designation of content rules represents the major difference between domain-referenced testing and objective-based evaluation. Content limits provide a set of rules to describe what content is appropriate to include or to sample in the test on instructional examples. The content limits describe the range of content to which the learner is expected to respond.

Referring again to the three components in the design of instruction (see p. 15), it can be seen that instruction is cyclical. That is, the three components are in constant feedback loops. Not only does the finished product get tested and revised, but even the objectives themselves are subject to revision. The result of such cycling is that the objectives, course content, and tests may eventually form parts of an interlocking system, where changes in one part will require adjustments in the other two parts. The advantage here is that the system can continue improving over a period of time.

Facilitates Student Learning. When behavioral objectives are given to students prior to the instructional content which is to be presented, they provide guidance to the student in the processing of information. Deterline (1968) says that if students are told precisely what the objectives are, in the form of minimum performance requirements, and if they are given sample test questions, performance can be improved. Behavioral objectives provide goals which are definable and assist in guiding the teacher in developing student activities. If behavioral objectives are used to tell the student exactly how he is going to be tested, the threatening aspect of the test will probably be reduced. Testing is a means by which students can check on their progress, or as a tool which the teacher uses to help them progress. By providing the objectives to the students one is communicating to the students what they are to do, to achieve satisfactorily. Kapfer (1970), too, advocates presenting behavioral objectives to students. He says that if students are given objectives they will be able to make intelligent choices concerning how they will attain them. In the past, students have not had this opportunity. Kibler et al. (1974) state that it seems reasonable that students who are presented with behavioral objectives are spared the frustration and time-consuming effort of trying to guess what the teacher expects of them. It also seems logical that students will learn more if they are told what is expected of them and how they will be expected to demonstrate that they have satisfactorily met the objective.

There are several more reasons for providing objectives to students (Duchastel and Merrill, 1974). The first is that behavioral objectives may provide direction to students' learning. Since they will know exactly what is expected of them, they will be able to discriminate between relevant and non-relevant material. The second is that objectives may provide some organization or general structure to the content or subject matter. Duchastel and Merrill also point out that objectives may serve a management function by enabling the students to better organize their time and learning experiences in terms of the goals of the course. This might help eliminate the typical cramming sessions which often precede tests. Another function is that of providing learners feedback in terms of the criteria set forth in the objective, enabling students to deal with any discrepancies between performance and goal. Finally, presenting objectives to students may help to motivate them. Students who know that they have satisfactorily met the criteria set forth in the objective will probably be more motivated than students whose only reinforcement comes from a grade at the end of a course. Duchastel and Merrill also point out that presenting objectives to students will have no results if the students pay no attention to them in the learning situation. Therefore, teachers must make an effort to thoroughly explain the meaning of objectives to students so that they will actually use them while learning. A discussion on the form and function of behavioral objectives would be helpful. However, teachers must be careful not to give long and extensive lists of objectives to students. This may overwhelm and confuse them. Such a list would defeat its own purpose.

Gagné et al. (1974) agree that the advantage of providing objectives to students is that it informs the learners of their goal. Gagné disagrees with those who contend that when one communicates an objective to students, they may be inhibited from trying to meet still other objectives which they may formulate themselves.

Informs the Teacher, Administrator, and the Lay Citizen of Purposes of Instructional Program. Accountability in education has gained acceptance from both the public and the federal government. With the growing involvement of parent groups in making decisions about local educational systems, it is clear that some form of accountability is needed. The public should be aware of the exact nature of what learning and schooling are all about. This type of accountability is becoming a more frequently discussed issue in education. Taxpayers, parents, funding agencies, and legislators are all extemely interested in having some type of proof that education in fact is taking place in our schools. Are schools really doing what they say they are doing? Why should any educator try to cover up what is being taught? How and what should "Johnny" know and do by the end of the school year? The answer to these questions is quite simple and straightforward: Specify the objectives, which in turn will inform these people about what we are doing and how we can prove it. To achieve the balance between spending and student learning that accountability demands, the teacher and school system must show evidence that students have learned as a result of their instruction. Educational accountability can be demonstrated successfully only when educational goals and objectives are precisely identified and stated. Kibler et al. (1974) say that the use of instructional bojectives will allow teachers to convey their goals to their supervisors and school boards. Burns (1972) suggests that specifying what is to be learned is obviously the function of objectives.

In order to defend budgets or requests for funds, administrators and teachers can provide the content of courses in objective form to the school board and thus demonstrate the need for expenditures to the board in more concrete terms. This process is much better than trying to provide verbal or verbal-pictorial representations of learning situations as they really exist, because board members are often too far removed from the classroom. Thus, instructional objectives may provide a basis for logical, concrete reasons for spending money. Scott (1974) states that objectives can also be used to explain to parents or to the community the philosophy on which a given course of instruction is based. Parents are often neglected in the educational process. However, parents are becoming increasingly concerned about the quality of education in the schools and are becoming more involved in the educational process. A list of objectives could be sent home to the parents telling them which objectives their child attained. Parents could then evaluate the progress of their child at intervals during the year and check to make sure that their child is keeping up. This would help inform parents about the content being taught as well as the childs' growth in the program. These objectives could also inform parents of the child's weaknesses and strengths. Such a procedure would be quite an improvement over the report care procedure commonly used.

There appear to be a least two advantages to the use of objectives in most teaching situations. First, objectives prompt teachers to determine the most significant aspects of the subject matter to be learned. The second is that objectives aid in establishing criteria for the measurement of classroom achievement. Since instructional objectives require teachers to specify criteria for acceptable behaviors and to determine in advance how satisfactory performance will be measured, teachers can achieve an increased sense of security. They feel more secure in their position and more satisfied with their professional contribution when they are confident in teaching the subject matter, confident of the subject matter's importance, and confident that the measurement techniques will measure whether or not the objectives have been met satisfactorily.

Instructional objectives are important at two levels of administration. The administrator who is in charge of curricula relies on objectives to insure that content and subject matter are covered adequately and that subject matter between courses does not overlap or become redundant. Instructional objectives also promote a thread of continuity among related courses. Instructional objectives developed by teachers give the supervising administrator insight into the teachers' philosophy and course goals. Teachers can collect data to determine the effectiveness of their instructional program and if students are continually failing to meet the standards set forth in the objectives, it may be a result of poor instruction. This in turn will enable administrators to more effectively evaluate teachers.

IV The Cases for and Against Behavioral Objectives

The Case for Behavioral Objectives. Since the time Mager's (1962) classic book on Preparing Instructional Objectives provided a major stimulus to the use of behavioral objectives in the field of education, two distinct schools of thought emerged, the first arguing the case for the use of behavioral objectives, and the other against the use. In arguing the case for the use of behavioral objectives in education, a large number of claims have been made. Proponents of the use of behavioral objectives maintain that behavioral objectives clearly indicate to students what is required of them, and as a result, student performance improves (Gagné, 1970, Mager, 1968, Popham et al., 1969, and Tyler, 1964). Objectives can also provide communication between the teacher and the student. Students become aware of where they are going and what is expected of them when objectives are given to them. Objectives work as an organizer. A considerable number of studies collected empirical data which indicate that giving objectives to students prior to instruction will enhance student learning. Dalis (1970) demonstrated that by using precise instructional objectives in advance of instruction enhanced learning of high school students in a health education class. The study implies, however, that objectives must be stated in precise terms, otherwise their value to the learning situation is doubtful. Doty (1968) investigated the effect of presenting objectives to students in a reading class. The results showed that the students who had prior knowledge of the objectives scored significantly higher on a posttest than did students who did not have prior knowledge. Lawrence (1970) studied two groups of students in a nursing care course, while Engel (1968) studied two groups of students in a mathematics course. Both studies reported that the group who received the behavioral objectives prior to instruction performed significantly better on a posttest. Blaney and McKie (1969) divided sixty volunteers into three groups, a behavioral objectives group, a general introduction group, and a pretest group. The results showed that the behavioral objectives group did significantly better than the introduction group on a posttest. The results also showed no significant difference between the pretest group and the behavioral objectives group on a posttest. Students in a college economics class were divided into two groups. One group received the behavioral objectives and the other did not. Tieman (1968) reports that by using retention scores as criterion, the behavioral objectives group scored significantly better than the non-objectives group.

There are also several studies that have shown no significant differences between groups of students who have received behavioral objectives and groups who have not (Boardman, 1970, Smith, 1967, and Weinberg, 1970). Therefore, the generalizability of providing objectives to students prior to instruction is not easily determined. The evidence reported here demonstrates the complexity of the issue. However, it has been shown that objectives sometimes help and are almost never harmful. Therefore, if the provisions of objectives are relatively inexpensive, one might as well make them available to students (Duchastel and Merrill, 1973).

Another reason for using behavioral objectives is that they serve as operational aids, basically because they are designed in terms of action (MacDonald-Ross, 1973). That is, they act as a medium of communication or a mechanism for informing people. Curriculum design is developed by the team approach quite often in our schools. By using well specified guidelines, in the form of behavioral objectives, each team member will know exactly what is being asked of him. Thus, the division of labor can become a much easier task.

A third claim for using behavioral objectives is that by specifying the exact behaviors one wants the students to exhibit, the teacher is better able to select appropriate learning activities or to design and suggest alternative instruction strategies appropriate to the individual learner. By constructing objectives that meet Mager's criteria for a well stated behavioral objective, the teachers will be guided in their choice of selecting instructional activities that will be specific, precise, and relevant to the desired outcome. Because of the systematic approach to instruction that behavioral objectives afford, the teacher can also pinpoint, at any time during instruction, those students who may be experiencing difficulty in achieving the objective. Thus, the teacher will be able to design and suggest alternative instruction strategies appropriate to the individual learner. In this sense, the use of behavioral objectives serves as an operational aid for the teacher in providing individual treatment for students. Because the outcomes of objectives can be replicated, treatment can be individualized. This means that students with different entry characteristics are recognized and remedial work can be provided for those who may need it. Tests developed on the basis of objectives, provide the teacher with diagnostic capabilities. MacDonald-Ross (1973) states that individualization may also mean that students can choose their own way to reach the objectives. For example, they might form contracts, which are written agreement between the teacher and student, to teach the goal or objectives specified by the teacher.

Clear and well sequenced objectives are necessary for individualization of instruction. Through testing, the teacher identifies where the child is academically at different times in the instructional program. Piper (1977) points out that frequent re-evaluation is needed in order to continue moving the child along in the instructional sequence. Careful and frequent recording of student progress is needed to facilitate the quality and accuracy of individualized instruction, as well as allowing for the evaluation and continued improvement of teaching techniques.

A fourth claim for using behavioral objectives is that they play a major part in the Objectives Based Instructional Programs that are being implemented in our schools. Classroom verified objective based programs have the potential for enabling teachers to provide students with the sufficient amounts of practice, feedback, and self correction needed to acquire competency on objectives is still very new to many teachers. Therefore, objective-based programs are assisting teachers in successfully promoting pupil attainment of the objectives by providing guidelines for effective teaching procedure.

However, the time that the teachers use to develop the instructional materials and procedures so that there is a high level of pupil achievement on the objectives requires extensive amounts of time, money, and expertise. Sullivan and Niedermeyer (1977) point out that if teachers want students to achieve mastery of objectives, teachers must also be provided with instructional materials and procedures developed especially for the objectives of the lesson. They go on to say that merely providing behavioral objectives to the teachers will have little effect on the learning situation. Teachers should not be accountable for high levels of pupil performance on objectives without the proper development of instructional materials and procedures. Sullivan and Niedermeyer conclude that with the increasing rate of objectives based programs in our schools, there is need for empirical evidence to support the concept of Objective Based Instructional Programs.

A fifth claim for using behavioral objectives is that they provide useful ifnormation for evaluating curriculum planning. Objectives guide the teacher and student in the teaching-learning process and they provide a measure against which progress can be judged. If teachers are to improve their teaching, they must have information that determines the success of different teaching methods and strategies that are used in instruction. Well-stated, clear objectives help provide this information. The information concerning curriculum planning evaluation is analyzed by using either norm-referenced or criterion-referenced testing procedures. Norm-referenced testing compares an individual's performance with that of a normative group. The standard in this type of testing is comparing a particular student's score with how other individuals performed on the test. On the other hand, criterion-referenced evaluation procedures are designed to determine whether a student has achieved mastery of a behavior as specified in an instructional objective(s). In criterion referenced testing the interpretation of a student's score performance is in no way dependent upon the performance of other students. Criterion-referenced procedures assume that if instructional objectives are important, teachers should be concerned with whether students have achieved them, not with how much they achieved relative to their peers. Objectives help to provide information on student achievement throughout the unit of instructiona as well as the end of the unit. Kibler et al. (1974) point out that criterion referenced testing is used for at least four different types of testing purposes: (1) for pre-assessment purposes, (2) for formative testing--to check on the progress of students so that assistance may be provided when necessary, (3) to determine whether components of instructional model need modification, and (4) to determine whether students have achieved the criterion levels of objectives at end of instructional unit. Teachers can determine whether students are ready to go on to higher order objectives or more complex objectives by evaluating lower level ones. Accurate assessment of what each student can and cannot do is critical for good teaching. It is a waste of time, as well as frustrating, to tell students what they already know about the content being presented or to present information that is "over their heads." Teachers have difficulty determining where their students are in the instructional process. With objectives clearly in mind, much of the guesswork is eliminated. When students see that they have achieved a satisfactory performance on lower level objectives, they are encouraged to further effort. Clearly stated objectives motivate both teachers and students.

Behavioral objectives may be informative in regard to curriculum planning. Bruton (1974) has done research on whether or not behavioral objectives would provide information for curriculum planning. The researcher evaluated a widely used set of oral language objectives from the Distar Language I Program (Englemann, Osborne, and Englemann, 1969) for use with first grade children. The results showed that fifty-four of the eighty-five objectives were shown to be useful in relation to instruction. The results also indicate that thirty-one of the objectives were not applicable for classroom use. Thus, the research cited shows that objectives can be used for guiding student learning.

The Case Against Use of Behavioral Objectives. Not all educators, by any means, favor the use of behavioral objectives. One concern is that of the origin of objectives--how are they derived? MacDonald-Ross (1973) states that there is no consistent view among educators as to the origin of objectives. He feels that two schools of thought have emerged concerning methods for deriving behavioral objectives. One group attempts to provide explicit rules for converting observable human action into behavioral objectives. MacDonald-Ross refers to this group as the "hardliners." They do not agree with the distinction between knowledge and skills and between education and training. The hardliners believe that one can observe a "master performer" at a task and be able to prescribe educational objectives. MacDonald-Ross feels that the task analysis procedure might be effective for skills but inadequate for general education. For example, trying to specify the objectives for a course in engineering by observing master performers would be quite fruitless unless you were also willing to take into account the network of knowledge and understanding underlying their actions. MacDonald-Ross states "the hardline case thus seems to fail. It is not sufficient to use observations of action (whether of action at work, or during examinations) for a prescription of educational objectives, if one takes the meaning of the word 'education' at all seriously."

The other group fully accepts that educational objectives need to be derived in a way which does justice to the difference between education and training. This group is constantly trying to justify the use of behavioral objectives in instruction. MacDonald-Ross (1973) refers to this group as the "softliners." Popham and Baker (1970) state that objectives are derived from three sources: (1) the learner, (2) the stociety, and (3) subject matter. They propose that philosophy of education and psychology of learning can help one formulate general objectives which can then be developed into precise instructional objectives. MacDonald-Ross feels that Popham and Baker's philosophy is frankly "hilarious" since it is far from being an operational procedure for deriving behavioral objectives. All the critical decisions seem to be left to intuition and common sense of the teacher. This does not reflect the concept of a systematic approach.

Many research studies in the area of behavioral objectives suffer from the lack of a precise operational definition of what a behavioral objective is. Duchastel and Merrill (1973) and MacDonald-Ross (1973) state that there is a lack of specificity in relation to determining whether objectives are behavioral or non-behavioral. Several studies indicated that some educators use Mager's three criteria for a well stated behavioral objective when defining the objectives, others did not. Some studies gave examples of well-stated objectives, while many others did not. This lack of operational definitions of the objective variables makes it difficult to compare studies.

The level of specificity needed in constructing behavioral objectives is another concern in the use of behavioral objectives. MacDonald-Ross (1973) states that there is an absence of rules for deciding what level of specificity objectives should be developed. General objectives have a tendency to become vague and ambiguous. In trying to eliminate ambiguity, one runs the risk of writing an impossibly long list of objectives. Wight (1973) states that when teachers begin writing objectives for what they hope to see as a result of instruction, they often write a behavioral objective for each intended outcome. This can result in a list so long that is more an obstacle course than an aid to learning for the student. Wight also states that meeting the specification for a properly written objective often inhibits productive thinking. Teachers get so caught up in the mechanics of writing behavioral objectives that they lose sight of their priorities, namely, what they are really trying to teach. MacDonald-Ross (1973) concludes that "every time behavioral objectives have been constructed on a large scale this problem of specificity has proved quite fearsomely difficult." He feels that no satisfactory rules have emerged and perhaps the problem of specificity is insolvable in principle. Woodruff and Kapfer (1972) state that the level of specificity in constructing objectives has seemed to transform education into a mechanistic program devoid of real human value. They point out that one of the most compelling critics of behavioral objectives, Charles Silverman (1970), states "Indeed, the approach to insturctional technology that most researchers are following (based on precise, measurable, behavioral terms) is likely to compound what is wrong with American education--its failure to develop sensitive autonomous, thinking, humane individuals" (p. 196).

Triviality is another criticism of behavioral objectives. In trying to satisfy the requirements of a properly written behavioral objective, teachers quite often find themselves listing inconsequential student behaviors and overlooking important ones. Meaningful objectives are more difficult to state in prescribed form. Pascal (1975) states that many educational goals which can be expressed in behavioral objective form are trivial, with the result that important outcomes of education are under-emphasized. Popham (1968) states that by going through the process of constructing explicitly stated objectives, educators can identify and weed out the trivial goals; then they can begin paying attention to more meaningful educational objectives. MacDonald-Ross (1973) states that while Popham's response that trivial objectives can be weeded out once revealed may be true, it still avoids the "huge" problems of origins and operationalism. He feels that the problem of triviality is still a problem.

One of the most fundamental problems with behavioral objectives is that the objective itself is confused with the indicator (means of determining whether the objective is achieved (Wight, 1973). A behavioral objective is a statement of a measurement to be taken, under specified conditions with criteria for evaluation to act as evidence that the desired behavior has been achieved. Many educators suggest that one should write a general statement or goal first. Then behavioral objectives can be written which relate to the general statements. But the general goal is often lost because of the focus on the behavioral objective; too much attention is devoted to the present performance specified in the indicator as opposed to the future capabilities called for in the general goal. Special effort should be made to relate the indicator to the goal. If this does not happen the student may find little meaning in the specification of performance.

Behavioral objectives frequently do not take into account that there may be many ways of assessing whether an objective has been achieved. This is particularly true when dealing with higher order cognitive objectives or objectives in the affective domain. An additional criticism of behavioral objectives is that they do not specify the measurement to be used. There are frequently many opportunities for assessing the behavior demonstrated within a learning activity and some of these may be better than those specified in the behavioral objective. It may be argued that behavioral objectives do not prevent educators from seeking other evidence to verify that the behavior has been achieved satisfactorily. But the point is that a well-stated behavioral objective has a measurement component that is explicitly stated, which causes the teacher not to look further for additional evidence that the objective has been achieved.

The question arises as to which behaviors or products of behavior can be accepted as valid for the purpose of objectives? Harlen (1972) states "behavior, by definition, is observable, but two observers would not always agree on the same interpretation of a particular item of behavior" (p. 226). Some behaviors may be observed only under cetain special condition. Are these described behaviors to be restricted to those objectives that can be observed and objectively judged only under certain condition? Harlen points out that observable behavioral changes are only sampled, and that indications that a change in behavior has taken place can never be a certainty. He feels that there is a blurred line between what is accepted as observable and what is not. Taba (1962) feels that deciding whether or not the criteria of the objective has been reached is less important. Harlen feels that the standard of acceptable performance in a behavior objective should not be rigidly applied and that some "slack" should be allowed, otherwise there is a potential drawback in specifying behavioral objectives with explicit criteria.

Some educators are expressing a distaste for the whole process of defining objectives. Sheehan (1974) states that advocates of behavioral objectives have become overly zealous and preach their message with a vengeance. Eisner (1967) states that many educators feel that the specification of objectives encourages students to seek the line of least resistance and thus lower their own educational goals. Other educators feel that there are important educational aims which cannot be translated into behavioral terms--attitudes, values and the creative experiences. Still others feel that the heavy emphasis on behavioral objectives implies training rather than education. Some educators believe that the behavioral objective movement will sterilize education. Sheehan (1974) states "The attempt to package, to circumscribe, and to modularize materials for the consumer is seen as counter-productive. It is in direct conflict with the more important goals of teaching students to identify their own educational requirements, to decide what they would like to learn and what is important to them within the limits of what is available and their own abilities." He goes on to say that the student must be able to learn on his own with the help of the following tools: books, films, television, journals, colleagues, his own observation, and the teachers. He must derive from them what is important without the aid of superimposed instructional objectives. Sheehan feels that behavioral objectives advocates have pushed their views too zealously and tend to see objectives as ends in themselves rather than as being only a small part in the instructional process.

A final criticism by educators in regard to use of behavioral objectives is that it takes an enormous amount of extra time and energy to formulate and use them. Conroy (1973) points out that most commercially produced instructional programs that do provide behavioral objectives describe outcomes that are usually limited to lower level skill capabilities. As a result, teachers are forced to write most of the objectives for the programs. Teachers feel that their role is becoming one of a clerk whose role is to mechanically grind out behavioral objectives. Conroy also feels that even though there is broad agreement about the benefits of using behavioral objectives in education, relatively little use is being made of the systematic process of developing instruction, and that rarely does a total school system manage its entire instructional program by behavioral objectives. He concludes: "The fact of the matter is that the vast majority of American teachers just do not use behavioral objectives as a part of their professional practice."

Conclusions

Although there are several logical reasons for using instructional objectives, there is limited empirical data to support their use. Kibler et al. (1974) point out that there are only about fifty or so experimental studies focused on instructional objectives. Unfortunately, the results of these studies are inconsistent. They also provide no conclusive evidence about the effect of instructional objectives on learning. Kibler states "of the thirty-three studies found that compared student learning with and without possession of instructional objectives." Current findings on the effects of instructional objectives provide no conclusive or consistent data on the relationship between the use of objectives and student learning. We are unable to draw any conclusive generalizations about the effect of behavioral objectives.

It was the purpose of this paper to present a review of the literature concerning the definition, form, function and the cases for an against the use of behavioral objectives. It is clear that there are many different definitions of the term "behavioral objective." Additional development is needed in this area so that a consistent operational definition of the term behavioral objective can for formulated. Although Mager's criteria for writing a well stated behavioral objective is the best known, literature reveals that there is little agreement as to which characteristic of the behavioral objective is most important, while others think that the standard or condition is. Gerlach et al. (1977) indicate that the choice of direct object also influences the observability and precision of a behavioral objective. They also contend that no single component is consistently of primary importance in determining the proper form of a behavioral objective.

Aside from the problem of determining the correct form of a behavioral objective and the need for training educators to write in an acceptable form, the literature suggests that educators must also be presented with the philosophy behind the concept of behavioral objectives. The basic characteristics of a behavioral objective are not fixed. Educators are not looking for objectives that are a particular size and shape, but are looking for objectives that are clearly stated and convey our instructional intents as concisely as possible.

Although there are disagreements regarding the definition and form of behavioral objectives, the literature reveals that behavioral objectives can have important functions in the instructional process: (1) aid in the design of instructional programs, (2) provide guidance in evaluation of instructional programs, (3) facilitate learning for students, and (4) inform teachers, administrators, and general public of the purposes of the instructional program. The major disagreements lie in the very nature of the behavioral objective movement itself. There is a wide disagreement as to the advantages of using behavioral objectives. MacDonald-Ross (1973) is one of the major critics, stating that "behavioral objectives will never achieve all that their supporter hope, for they are limited by the very presumptions on which they are based."

As an advocate for the use of behavioral objectives in education, I will conclude with the following impressions that I have drawn from review of the literature.

1. There is a need to generate explicit principles for constructing relevant behavioral objectives. Rules for specificity of objectives need to be developed.
2. Behavioral objectives form a well-worked out method of rational planning in education. They encourage educators to make explicit the values they may have never revealed as well as making them think and plan in detailed, specific terms.
3. Behavioral objectives help to better organize the students' time as well as give them direction and provide motivation.
4. Behavioral objectives can form the basis for a well-worked out program for individualizing instruction.
5. Behavioral objectives are the clearest verbal devices available to educators for use in communicating the intent of learning programs to students, administrators, and general public.
6. Behavioral objectives can provide direction and guidance to teachers when they are choosing instructional activities and materials for a learning program.
7. Behavioral objectives provide a rational basis for evaluating instructional programs. Even with the high "costs" attached, the cyclical approach to evaluating instructional programs is worthwhile. Good instruction is not developed overnight and without expense.

References

Baker, E. L. Beyond objectives: Domain-referenced tests for evaluation and instructional improvement. Educational Technology, 14(6): 10-16 (1974).

Barron, M. C., Gerlach, V. S., & Haygood, R. C. An Empirical Analysis of the Components of Behavioral Objectives (Technical Report No. 60331). Tempe, Arizona: Arizona State University, 1976.

Blaney, J. P., & McKie, D. Knowledge of conference objectives and effect upon learning. Adult Education Journal, 29: 98-105 (1969).

Bloom, B. S., Englehart, M. D., Furst, E. S., Hill, W. H. & Krathwahal, D. R. Taxonomy of Educational Objectives: Handbook 1, Cognitive Domain. New York: David McKay Company, Inc., 1956.

Bloom, B. S. Learning for mastery. In B. J. Bloom, J. T. Hasting, & G. F. Madaus (Eds.), Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill, 1971.

Boardman, E. The effects of advanced knowledge of behavioral objectives on students' achievement in remedial chemistry. Unpublished doctoral dissertation, University of Southern California, 1970.

Briggs, L. J. Handbook of Procedures for the Design of Instruction. American Institute for Research, Washington, D. C., 1970.

Bruton, R. W. A method for evaluating instructional objectives. Educational Technology, 19(9): 33-38 (1974).

Burns, R. W. Behavioral objectives for competency-based education. Educational Technology, November, 12(11): 22-25 (1972).

Conroy, W. G. The synthesized behavioral objective. Educational Technology, 13(10: 52-58 (1973).

Dalis, G. T. Effort of precise objectives upon student achievement in health education. The Journal of Experimental Education, 40(5): 262 (1970).

Deterline, W. A. The secrets we keep from students. Educational Technology, 8(2): 7-10 (1968).

Deno, S. L. & Jenkins, J. R. On the behaviorality of behavioral objectives. Psychology in the Schools, 6, 18-24 (1969).

Doty, C. R. The effect of practice and prior knowledge of educational objectives on performance. Unpublished doctoral disseration, Ohio State University, 1968.

Duchastel, P. C. & Merrill, P. F. The effects of behavioral objectives on learning: A review of empirical studies. Review of Educational Research, 43(1): 53-59 (1972).

Eisner, E. W. Educational objectives: Help or hinderance. School Review, 75(3): 250-260 (1967).

Engel, R. S. An experimental study of the effect of stated behavioral objectives on numeration. Unpublished Master's thesis, University of Maryland, 1968.

Engelmann, S., Osborne, J., & Engelmann, T. Distar language 1: An instructional system. Chicago: Science Research Associates, 1969.

Esbensen, T. Writing instructional objectives. Phi Delta Kappan, January, 48, 246-247 (1967).

Gagne, R. M. The Conditions of Learning (2nd ed.). Holt, Rinehart and Winston, 1970.

Gagne, R. M. & Briggs, L. J. Principles of Instructional Design. New York: Holt, Rinehart and Winston, 1974.

Gerlach, V. S. On the behaviorality of behavioral objectives--A replication (Technical Report No. 40701). Tempe, Arizona: Arizona State University, 1974.

Gerlach, V. S., & Barron, M. C. What makes a behavioral objective "behavioral" (Technical Report No. 41231) Tempe, Arizona: Arizona State University, 1974.

Gronlund, N. E. Stating Behavioral Objectives for Classroom Instruction. Toronto, Canada: Collier-MacMillan, 1970.

Harlen, W. Formulating objectives-Problems and approaches. British Journal of Educational Technology, 3(3): (1972).

Harrow, A. J. The behavioral objectives movement: Its impact on physical education. Educational Technology, 17(6): 31-39 (1977).

Haygood, R. C., Gerlach, V. S., & Wigand, D. L. The effects of four components of behavioral objectives of raters' perceptions (Technical Report No. 70228). Tempe, Arizona: Arizona State University, 1977.

Hempel, C. G. The theoretician's dilemma: A study of the logic of theory construction. Aspects of Scientific Explanation, Glencoe, Illinois: Free Press, 1965.

Kapfer, P. G. Behavioral objectives and the curriculum processor. Educational Technology, 10(5): 14-17 (1970).

Kapfer, P. P. Expanding behavioral objectives for educational design. Educational Technology, 17(5): 16-21 (1977).

Kibler, R. J., Cegala, D. J., Barker, L. L. & Miles, D. T. Objectives for Instruction and Evaluation. Boston, Massachusetts: Allyn and Bacon, 1974.

Kibler, R. J., Barker, L. L., & Miles, D. T. *Behavioral Objectives and Instruction*. Boston, Massachusetts: Allyn and Bacon, 1974.

Krathwohl, D. R. Stating objectives appropriately for program, for curriculum and for instructional materials development. *Journal of Teacher Education*, *16*, 83-92 (1965).

Lawrence, R. M. The effects of three types of organizing devices on academic achievement. Unpublished doctoral dissertation, University of Maryland, 1970.

Levinson, H. Management by objectives. *Harvard Business Review*, 48(7): 32-38 (1970).

Lindvall, C. M. *Testing and Evaluation: An Introduction*. New York: Harcourt, Brace & World, 1961.

MacDonald-Ross, M. Behavioral objectives--A critical review. *Instructional Science*, 2(1): 1-52 (1973).

Mager, R. F. *Preparing Instructional Objectives*. San Francisco: Fearon Publishers, 1962.

Mager, R. F. *Developing Attitudes Toward Learning*. Belmont, California: Lear Siegler, Fearon, 1968.

Merrill, M. D. Necessary psychological conditions for defining instructional outcomes. *Educational Technology*, 11(8): 34-39 (1971).

Niedermeyer, F. C. & Sullivan, H. J. Prospects for school acceptance of objectives-based instructional programs. *Educational Technology*, 17(6): 20-26 (1977).

Palmer, R. H. 200 years BB (before Bloom). *Educational Technology*, 14(1): 24-26 (1974).

Pascal, C. E. Towards meaningful educational objectives. Improving College and University Teaching, 23(3): 145-148 (1975).

Piper, T. A synergistic view of behavioral objectives and behavior modification. Educational Technology, 17(6): 26-31 (1977).

Popham, W. J. Probing the validity of arguments against behavioral goals. Paper presented at the American Educational Research Association Convention, 1968.

Popham, W. J. & Baker, E. L. Establishing Educational Goals. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.

Scott, H. V. Levels of involvement: A descriptive model for teacher education. Educational Technology, 14(3): 48-52 (1974).

Sheehan, T. J. Why the cool reception to behavioral objectives in the university? Educational Technology, 14(11): 54-56 (1974).

Silberman, C. E. Crisis in the Classroom. New York: Random House, 1970.

Smith, S. A. The effects of two variables on the achievement of slow learners on a unit in mathematics. Unpublished Master's thesis, University of Maryland, 1967.

Smith, P. G. On the logic of behavioral objectives. Phi Delta Kappan, 53, 429-431 (1972).

Sullivan, H. J. Objectives, evaluation, and improved learner achievement. In Popham, W. J., Eisner, E. W., Sullivan, H. J., and Tyler, L. L. Instructional Objectives, AERA Monograph on Curriculum Evaluation, No. 3. Chciago: Rand McNally, 1969.

Taba, H. Curriculum Development. New York: Harcourt, Brace & World, 1962.

Tieman, P. W. Student use of behaviorally-stated objectives to augment conventional and programmed revisions of televised college economics lecturer. Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1968.

Tieman, P. W. Analysis and the derivation of valid objectives. NSPI Journal, July, 1969, 8, 16-18.

Tieman, P. W. Conceptual objectives for domain-referenced education. Educational Technology, 17(5): 21-22 (1977).

Tuckman, B. W. Conducting Educational Research. New York: Harcourt, Brace and Jovanovich, 1972.

Tyler, R. W. Constructing Achievement Tests. Columbus, Ohio: The Ohio State University Press, 1934.

Tyler, R. W. Some persistent questions on the defining of objectives, in Lindvall, C. M. (Ed.). Defining Educational Objectives, Pittsburgh, Pennsylvania: University Press, 1964.

Weinberg, H. Effects of presenting varying specificity of course objectives to students on learning motor skills and associated cognitive material. Unpublished doctoral dissertation, Tempe, Arizona: Arizona State University, 1970.

Wight, A. Behavioral objectives in Perspective. SRIS, The Quarterly Publication of Phi Delta Kappa's School Research Information Service, Winter, 1973.

Woodruff, A. D., & Kapfer, P. G. Behavioral objectives and humanism in education: A question of specificity. Educational Technology, 12, 51-55 (1972).

RELATING BEHAVIORAL SCIENCE RESEARCH TO PRACTICE:

WHAT, WHY, HOW

Dr. Malcolm L. Fleming
Audio-Visual Center
Indiana University

Paper presented at Annual Meeting, Association for Educational Communications and Technology, New Orleans, March, 1979

Abstract

The troubling problem of relating research to practice is discussed with reference to: what, why, how. The what of recent research is illustrated by examples which suggest media-related research is becoming more rigorous, analytical, and theoretical, and thus more reliable and generally applicable. The why is briefly argued with reference to the greater reliability of research-based information over that learned on the job by an individual practitioner. The how is considered by noting the large available reservoir of relevant findings and by describing a way of cognitively relating the practitioner's questions to the researcher's findings through their common interest in relating instructional conditions to instructional effects.

RELATING BEHAVIORAL SCIENCE RESEARCH TO PRACTICE:

WHAT, WHY, HOW

Nine years ago I reported to this group on some perceptual principles for the design of instructional materials. The occasion was the meeting in Detroit of what was then DAVI. I stressed five principles based on perception research: Perception is very selective, is influenced by a prior set to see some things and not others, and is highly organized. Further, auditory perception is specialized for temporal distinctions and visual for spatial distinctions, and finally, pictures of objects are better remembered than their names. Since then Dr. Levie and I have searched out numerous other principles from behavioral science research: 56 from the perception literature, 41 from the memory literature, 27 from the concept learning literature, and 69 from the attitude change literature. These are now packaged in a book entitled Instructional Message Design (1978). Consequently, one of the things this paper will not include is a reiteration of any of those research-based principles for the design of instructional materials. Instead, I will address the general problem of moving research into practice and do so from three perspectives: the what, the why, and the how.

First, the what. Obviously it is impossible in these few minutes to deal adequately with what behavioral science research is being reported today and what part of that is particularly pertinent to instructional designers, developers, and media users. So, I will briefly describe three trends in media-related research and give a few examples of each. Research today differs from that of twenty years ago in several important ways: it is more rigorous, more analytical, and more theoretical. Although these characteristics of research may fail to describe the applied research which many professional practitioners identify with, I will argue and give examples to support the opposite view, i.e. that it is the rigorous, analytic, and theoretical research that best serves the practitioners needs.

Trend number one, *more rigor*. The idea of rigor is simply that research is better conceptualized and better controlled, and hence more reliable. This frequently takes the form of replicating prior studies or re-addressing previously studied problems with improved procedures. For example, there was a long-accepted principle that knowledge of correct responses (KCR) would be reinforcing to the learner and hence would be associated with increased learning. This was also a basic tenet ~~tenant~~ of programed instruction (PI), until study after study failed to support it. A study by Anderson, Kulhavy and Andre (1971) took a more rigorous approach. These investigators noted that reinforcement (KCR) was by definition, to *follow* the correct response and, further that the response was to be made to the *criterial* cues. Most prior PI studies had not controlled for these factors. Subjects could peak at the correct response *before* responding and could readily do so without paying much attention to the criterial cues. Consequently, conditions were arranged, by way of computer terminals, such that subjects were required to respond before receiving KCR and were variously encouraged to give careful attention to criterial cues. The results, for once, showed a significant difference in recall favoring the KCR condition over the no KCR. This seemed to resolve the issue, but Anderson and Kulhavy went further, they replicated the study but included both Peak and

No-peak conditions in a controlled comparison. The "Peak" group made higher within-program scores, but the No-Peak group made significantly superior recall scores after the program. Increased research rigor, thus, produced more reliable findings which are of direct relevance to practitioners.

In another series of studies, Levin and his associates had been studying the effects of two treatments on children's learning of prose stories. Repeatedly, the prose-plus-related-pictures condition was superior to the prose alone, the stories being read aloud in both cases. For investigators interested in picture effects, these results were quite positive, but rigor required that they consider plausible alternative interpretations of these data. One alternative was that the pictures contributed nothing uniquely attributable to their being pictures, for they might be serving only as a repetition of the story, which repetition could as well have been provided verbally. A follow-up study (Levin, 1976) included a verbal repetition condition in which each sentence was repeated aloud either by the narrator or by the learner. Results showed that verbal repetition increased learning over no-repetition, but the single verbal plus pictures was still significantly superior to all other conditions.

A more general sign of greater rigor in our field is the marked reduction in the number of gross media comparisons, for such studies are generally both conceptually weak and procedurally very difficult to control.

Trend number two, more analytical. The most apparent indication of increasingly analytical research is the greater numbers of independent variables examined in a study, i.e. not just media variables but also learner variables and subject matter or task variables. There is also an increase in the diversity of dependent variables, i.e., not just recall, but also recognition, concept learning, problem solving, attitude change, and physiological changes. Regarding learner variables, for example, a greater diversity of attributes are being examined, e.g., sex, cognitive style, development level. Results include much evidence that females are generally more skilled in verbal tasks than males, while there is some evidence that males are generally more skilled in some spatial tasks than females.

People also differ in important ways called cognitive style. For example, some are more skilled than others at distinguishing a simple visual figure from a complex background. A surprisingly large amount of research has been done with this basic visual perception skill, some of it demonstrating a relationship to other cognitive skills and even to social skills. This so-called cognitive style is not related to general intelligence but does seem related to the type of school subject and type of instruction method that learners can most readily cope with. Subjects who are most skilled at separating figures from complex grounds (called field-independents) tend to be more analytical, prefer science courses and and working alone, and are better able to organize information themselves. Subjects least skilled at figure-ground separation (called field-dependents) tend to be cognitively more global, prefer social studies courses and working with people, and learn best from well organized learning materials (Witkin, et al., 1979). These more analytical findings all have interesting implications for instructional developers.

Pictures are also being studied more analytically. For example, Margaret Hagen (1974) has been studying how we perceive pictures and to what extent and in what ways we must learn to perceive them. The literature is contradictory, but it does appear that infants can recognize without training a picture of a familiar object. However, skill in perceiving certain aspects of complex pictures must apparently be learned. For example, we must learn to adjust to impoverished information e.g. lack of certain depth cues in flat pictures, and lack of motion cues in still pictures, and must learn to interrelate various shades, textures, and forms so as to reduce their intrinsic ambiguity. Jean Mandler (1976) has studied the kinds of information which people selectively remember in pictures. Evidence to date suggests that the identity of objects in a picture is remembered best, the spatial location of those objects somewhat less well, and the interior details of the objects least well. Needless to say, these analytical studies are highly relevant to professional practice in our field.

Analytical studies produce quantities of findings which are complexly related to numerous variables. The effect is to greatly complicate the task of interrelating such studies so as to summarize, generalize and form some useful conclusions. It is here that overriding conceptualizations, models, and theories are indispensable.

Trend number three, more theory. Theories are not only more general than individual findings, they also explain why, and thus are of great potential utility.

Although numerous theories have been generated, it is still not clear why pictures are more memorable than words, but the effort to explain this general finding has produced much important research. For example, several conditions have been identified in which pictures are not superior, e.g., where the pictures are presented at a rapid rate and also where they are quite similar to each other.

One of the best supported theories is dual-coding, i.e., that pictures are more memorable because they are more likely to be coded twice, both imaginally and verbally, and dually stored. Words are less memorable because they are less likely to be dually coded and stored. However, it is clear that both are readily recoded, i.e., we see a picture of a friend and think of his name or we hear or read his name and think of a mental picture of his face. The consequences of this for media professionals are potentially shattering, for the cognitive effects of our stimuli are not given by their objective form--the effect of a concrete picture may be an abstract verbal concept. Like it or not, these and other theoretical insights will change our professional practice.

Dual-coding theory is similar to others in which explanations are sought at the processing level through such questions as: How are external stimuli represented internally, how are they changed/translated/elaborated during such processing, how are they stored in memory, and how are they retrieved ~~retreived~~ and re-presented to awareness?

A contemporary theoretical account is called depth-of-processing, i.e., the greater the depth (or breadth) of processing of a picture or word the greater the recall. For example, Bower and Karlin (1974) found significant differences in recognition memory for pictures of people depending on the depth of processing presumably induced. Subjects instructed only to judge the sex of the pictured people remembered the pictures less well than those instructed to judge the honesty or the likeableness of the people. One of the implications for message designers is that the mere presentation of pictures is not enough. The learner must be further involved in studying them (by way of introduction or task), in attending to the most critical aspects (by way of pointers, captions), in thinking about them (by way of answering questions, solving problems). This theoretical research thus yields another practical insight: the learner must process information in more than superficial ways, and can be influenced to do so by appropriate instructional design.

Theoretically oriented research on visual processing has produced many very interesting and relevant insights for our field. The mental representation of a picture, object or event, i.e. a picture in the mind, has been shown to participate in numerous cognitive processes. A mental image can markedly facilitate memory for both words and pictures (Paivio, 1971); it appears to contribute to concept formation by representing the most common or prototypic examples of the concept (Rosch, 1978); it appears to function importantly in the solving of spatial problems and of spatial games like chess (Chase and Simon, 1973); it is widely reported to have been the medium of creative and inventive thought used by both artists and scientists, including Einstein, (Shepard, 1978). There is also evidence that thinking visually is a cognitive skill that can be developed through repeated experience with visual media such as film (Salomon, 1978). The implications of such theoretical work on mental imagery are mind boggling, especially for visual media researchers and practitioners.

In sum, I have endeavored to indicate that media-related research has become increasingly rigorous, analytical, and theoretical, and that the effect for professional practitioners has been both an increased reliability and a broader generality of conclusions. So much for the what of relating research to practice.

The *why* of relating research to practice may need no discussion for this group, it being understood that a profession is characterized by the reliability of its knowledge and the relevance of that knowledge to its practice. If our research does not inform our practice there is little justification for our profession to engage in it. However, one can ask, "Why bother relating research to practice where a seasoned professional will have discovered many of the answers simply by facing a variety of problems, trying a variety of solutions, and observing the results?" One argument for research even under these circumstances is that the practitioner primarily learns from his own individual experience while the researcher generally builds upon the accumulated knowledge of numerous other investigators. Further, researchers generally are more precise in setting up the "test" situation and in analyzing and evaluating the results; and they more regularly go public with their results so that others can criticize them and develop other studies to repeat or contradict the findings. For these several reasons research tends to produce more systematic and reliable information than practice, at least in those domains of human experience that are amenable to research methodology. But in addition to reliability, research must also pass the test of relevance before it can be successfully related to practice. It is here that my third topic, the *how* of it, comes in. *How* can research be made relevant to practice?

My answer assumes that the research in question has been or will be well done, i.e. meets the criteria previously described: rigorous, analytical, and theoretical. My answer also assumes that the research is potentially useful to media practitioners, and that includes a surprisingly large proportion of the very extensive body of research on instruction, learning, and communication. Often a practitioner making a film on earth science for eighth graders, for example, feels that the only relevant research is that on films for eighth grade science. Actually many areas of research may be pertinent: how eighth graders learn in general, the effect of their stage of cognitive development, and their interests; the characteristics of science as a form of knowledge, the differences between learning facts, concepts, and skills; the potential strengths and weaknesses of film in instruction, the fixed pace, the lack of learner control. Also pertinent is the research on the many stimulus elements available in film such as motion, color, pictures vs. words, pictures interacting with words, the sequence or order of pictures and of words, the characteristics of visual perceptual channel vs. auditory channel, etc., etc. The extensive research on all these factors assures that there is no dearth of potentially relevant research findings for the earth science film maker or most any other instructional developer.

Professional practice in our field is frequently represented by a model which depicts a series of steps or operations which the practitioner must perform or supervise. Research is seldom depicted in these procedural models, though it has sometimes been suggested as somehow underlying the procedures or surrounding them, the idea being that practice somehow takes place in a nutrient medium of research. But the general absence of reference to research in development models must be explained in some way, e.g., research is considered irrelevant or only remotely related, or research is assumed throughout and hence not depicted, or knowledge of research (as with other forms of knowledge) doesn't belong in an intrinsically procedural model. I prefer the latter interpretation, that research represents an essential but non-procedural aspect of the process.

While there is important research on development procedures themselves, it is my view that most current research pertinent to design and development impinges on the way in which professionals conceptualize or think about the process. What kinds of thought processes are most productive of viable solutions to design and development decisions? Davies (1971) and others have written about the process of decision making. My concern is with the substance, the content, of the decision-makers thought processes, for it is here that research converges on practice.

Probably the question most central to the developer's thoughts is: What instructional conditions will most effectively and efficiently guide the learner to the competence desired? This reduces to the basic relation between instructional conditions and learning effects, or more fundamentally to the relation between conditions and effects. Both research and practice can identify with this relationship. The researcher's interest is in studying the relation between conditions (the independent variables) and effects (the dependent variables), while the practitioner's interest is in deciding which condition will most likely produce the effect he desires. It is at this commonly held basic conceptual level, relating conditions to effects, that I believe the rapprochement between research and practice can most appropriately be made. The commonality that makes rapprochement possible can be represented by the parallel forms of two statements: one the researcher's conclusions and the other the developer's questions. Researchers conclude:

Certain conditions have lead to certain effects.

Developers ask:

What conditions will lead to the desired effects?

Clearly the juxtaposition of the developer's questions with the appropriate researcher's conclusions has the effect of relating research to practice in a direct way. A designer/developer who conceptualizes, thinks about, and discusses decisions in this way is more likely to make reliable decisions, particularly if he has in mind an adequate supply of reliable statements linking conditions to effects. These linking statements I call instructional design/development principles. Examples are the perception principles cited at this convention almost ten years ago and the numerous others cited since.

Lest anyone think this cognitive matching of research findings and practitioner questions is so pat a process as to automatically produce a correct solution must realize that the principles do not apply to all design/development problems and even where applicable serve only as fairly reliable guides. They must always be tested in each design/development application, i.e., prototypic messages must be given careful formative evaluation.

In sum I have endeavored to report a few examples of recent research in our field that are increasingly rigorous, analytical, and theoretical, and thereby are more reliable and generally useful for professional instructional designers and developers. Further I have provided arguments why much research in instruction, learning, and communication is relevant to our professional practice and have described a cognitive process model for relating research findings to designer/developers questions.

References

Anderson, R.C., Kulhavy, R. W., & Andre, T. Feedback procedures in programmed instruction. *Journal of Educational Psychology*, 1971, *62* (2), 148-156

Bower, G.H., Karlin, M. B., Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, 1974, *103*(4), 751-757

Chase, W.G., Simon, H. A. The mind's eye in chess, in W. G. Chase, Ed., *Visual Information Processing*. New York: Academic Press, 1973

Davies, K. K. *The Management of Learning*. London: Mc Graw-Hill, 1971

Fleming, M. L., Levie, W. H. *Instructional Message Design: Principles from the Behavioral Sciences*. Engelwood Cliffs, New Jersey: Educational Technology Publishers, 1978

Hagen, M. A. Picture perception: Toward a theoretical model. *Psychological Bulletin*, 1974, *81*(8), 471-497

Levin, J. R., Bender, B. G. & Lesgold, A. M. Pictures, repetition, and young children's oral prose learning. *AV Communication Review*, 1976, *24*, 367-380

Mandler, J. M., Johnson, N. S. Some of the thousand words a picture is worth. *Journal of Experimental Psychology: Human Learning & Memory*. 1976, 2(5), 529-540

Paivio, A. *Imagery & Verbal Processes*. New York: Holt Rinehart & Winston, 1971

Rosch, E. Principles of categorization, in E. Rosh & B. B. Lloyd, Eds., *Cognition & Categorization*, Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1978

Salomon, G. The "language" of media and the cultivation of mental skills. Paper presented at the Annual Meeting of American Educational Research Association, Toronto, 1978

Shepard, R. N. The mental image. American Psychologist, 1978, 23, 125-137

Witkin, H. A., Moore, C. A., Goodenough, D. R., & Cox, P. W. Field-dependent and field-independent cognitive styles and their educational implications. Review of Educational Research, 1977, 47(1), 1-64

APPLYING THE ATI CONCEPT IN AN OPERATIONAL ENVIRONMENT

by

Perrin E. Parkhurst, Ph.D.
Michigan State University[1]

IR001194

Paper presented at the Annual Meeting of the Association for Educational Communications and Technology, New Orleans, LA, March 1979.

---

[1] Dr. Parkhurst was formally with McDonnell Douglas Corporation, Advanced Instructional System (AIS), Denver, Colorado.

ABSTRACT

This paper describes a model for the systematic design, development, and validation of alternative instructional modules, by incorporating the basic concepts underlying ATI analyses. The proposed working model is based on the premise that the ATI concept is best applied, not in isolation, but as an integral part of a dynamic decision making instructional environment. Four areas of interest, for applying the ATI concept in an operational environment, are discussed: (1) Operational Methodology; (3) Variable Categories; (3) Statistical Procedures; and (4) Alternative Module Development Guidelines.

APPLYING THE ATI CONCEPT IN AN OPERATIONAL ENVIRONMENT

Overview

For many years, much has been written about ATI methodology, ordinal and disordinal interactions, and the search for meaningful hypotheses involved with ATI research. In this brief paper, the writer has attempted to synthesize both prior work experience and prior readings into a working model. This model attempts to set forth a framework for designing alternative instructional modules that have rewards for both the learner and instructor.

An aptitude variable can be any personological or organismic variable upon which individuals differ (e.g. IQ, anxiety, dogmatism, etc.). A treatment is any instructional strategy or combination of instructional strategies that structures information for the purpose of having students learn that information. As generally defined, an aptitude-treatment-interaction exists when, as a result of a given treatment, individuals at one end of an aptitude variable perform at one level on a criterion measure. Individuals at the other end of the aptitude variable perform at a significantly different level on the criterion measure and the reverse trend holds true for a second treatment. (For an expanded discussion, see Snow and Salomon, 1968; Bracht, 1970; Rhetts, 1974; Snow, 1977; Cronbach and Snow, 1977.)

The literature is filled with plots (varying greatly in degree of angularity) of different treatment groups against both aptitude and criterion measures. However, little if anything is written which proposes how to incorporate the ATI concept into a decision making instructional environment. Two common criticisms of the ATI approach are that generalizations resulting from ATI research: (1) lack a design strategy that demonstrates a need to develop alternative modules in the first place; and/or (2) are not easily interpreted so as to provide detailed guidelines for instructional designers. The proposed working model is based on the premise that the ATI concept is best applied, not in isolation, but as an integral part of the dynamic decision making instructional environment.

Four areas of consideration, for applying the ATI concept in an operational environment, will be presented: (1) Operational Methodology (overall master plan) (2) Variable Categories (for independent and dependent variables), (3) Statistical Procedures (analyzing the results), and (4) Alternative Module Development Guidelines (detailed guidelines for instructional designers).

Operational Methodology

Underlying the implementation of this model is the assumption that for most of us, the time and expense involved in alternative module development is not worth taking unless the existing instructional module causes large or alarming variations in the criterion variable. That is, if everyone is reaching criterion or achieves mastery (passing grade) in an acceptable time frame within a given module, why use resources searching for ATI's or developing alternate modules? It is only when grades, time to completion, or some other dependent variable of interest and importance varies within a given module, that educators should be concerned about developing alternate modules. Only then does it seem justifiable to allocate resources for alternative module development.

The operational methodology suggested here involves five main steps (refer to Figure 1):

Step 1: Establish Main Track Module(s) (Point "A")

Establish a single or main track of instruction; that is, develop a "best guess" approach by designing one module per instructional concept for the entire unit or course being taught.

Step 2: Establish Data Analysis Pool (Point "B")

Establish a data pool of variables of interest and importance. These variables must be ones that are expected to be closely associated with performance on the criterion measure.

Step 3: Assess Variance (Point "C")

Examine the dependent variable of interest, usually time or score, with regard to amount of variance in main track instruction. Identify main track modules with unacceptable variance in the dependent variable.

Step 4: Analyze Data (Points "D" and "E")

Analyze data and, based on significant predictor set(s), develop alternative modules for main track modules, only where large variations exist in the dependent variable.

Step 5: Validate Alternative Module(s) (Points "F" and "G")

Validate alternative module development. Examine actual performance data associated with predicted performance data under varying module assignment rules; i.e., random vs. non-random for both main track and alternative module(s). Assess results and make decision(s).

Variable Categories

Once the main track modules(s) have been developed (Step 1,) serious consideration must be given to the selection of variables to be entered into the data pool. At least four categories of independent variables are suggested for consideration during Step 2 (Figure 1, Point "B"):

1. Cognitive
2. Achievement Motivation
3. Personality/Anxiety
4. Derived

COGNITIVE - This would be either a general mental ability measure or an aspect of general mental ability for which a strong rationale could be built that this aspect is related to performance on the dependent variable (e.g. high spatial ability skills required: measured by performance on a visual discrimination test).

ACHIEVEMENT MOTIVATION - This element of the data pool is extremely important in determining the degree to which an individual's motivation to succeed is related to the task at hand. This variable must be designed to measure specific task achievement motivation for the dependent variable of interest (e.g. high success drive required: measured by performance on a test to identify potential corporate executives). For an excellent discussion of factors to consider in the selection of this type of variable, see deCHARMS and MUIR (1978).

PERSONALITY/ANXIETY - According to Cronbach and Snow (1977) analysis of achievement via levels of anxiety is a potentially useful concept that has emerged from analyzing many ATI studies. The optimal level between anxiety and various levels of state/trait curiosity as predictors for differing kinds of performance tasks has yet to be widely studied, much less reported in detail. Among those who have written on this topic are: Leherissey, O'Neil, and Hansen (1972 and 1973); Spielberger (1972); and McCombs (in press).

DERIVED - A fourth category of independent variable that may deserve attention in more sophisticated instructional environments is a category often called a "within-course" or "derived" variable. It is simply using time on one module (or some other variable whose value is not determined until after the start of an instructional unit) to predict time/score (or some other dependent variable of interest) on subsequent modules. This set of variables is often only able to be studied in very sophisticated computer managed systems where dynamic decision-making and data collection are possible.

Dependent variable(s) must also be selected. These are usually either time to completion, number of trials to completion, total score, or some combination of all three.

Statistical Procedures

There are basically two stages in the data analysis procedures. They involve:

1. Identifying the main track modules with large variations in the dependent variable (Figure 1, Point "C").
2. Using stepwise multiple regression to identify a predictor set that is predictive of dependent variable values (Figure 1, Point "D").

ASSESSING VARIATIONS - Admittedly identifying main track module(s) with excessive variations in the dependent variable is often an arbitrary decision. There are many sophisticated statistical techniques to determine empirically what constitutes large or significant. However, the emphasis here is on presenting a model or procedure that stresses operational considerations.

Preliminary examination simply of range, standard deviation and mean score for each main track instructional module dependent variable can provide a wealth of usable information. Each educator must then make some hard decisions about what is acceptable performance variability. When modules have unacceptable or large variations in performance, they then become candidates for further and more discriminating statistical analysis. Where large variations in main track modules do not exist, it only makes good sense to either continue using that module or make minor revisions and repeat the analysis process.

DETERMINING PREDICTOR(S) - Once a module has been identified as having unacceptable variations in the dependent variable (that is, some sub-group of students performed well or quickly while another sub-group of students performed poor or slowly), the task of determining the best predictors of performance begins.

The analysis procedure suggested here is to use stepwise multiple regression to identify, from the previously obtained data analysis pool, significant predictor set(s) for the performance variable. By examining the complex nature of significant predictors, a more accurate estimate of those variables specifically associated with performance by a sub-group of students, when attempting a given task, is possible. Lacking a theoretical framework for the selection of variables to be entered into the analysis pool would be a major deficiency in the implementation of the procedures suggested here.

Alternative Module Development Guidelines

Once the predictor set is identified, alternative module development begins (Figure 1, Point "E"). For illustration, assume the following hypothetical situation exists when attempting to predict a score on a particular main track module.

Predictors	Simple r	Mult R	R^2
Curiosity	+.40	.40	.16
Spatial Ability	+.38	.56	.31
Verbal Reasoning	-.35	.72	.52

In this situation, we see that curiosity and spatial ability are positively related to performance while verbal reasoning is negatively related; also, 52 percent of the variance in the dependent variable is accounted for by combining all three independent variables as predictors. One instructional strategy that might be worth pursuing would involve building an alternative module that aroused curiosity while compensating for lack of spatial ability. This alternative module would be hypothesized to be appropriate only for low verbal reasoners. Such a module might have advanced organizers that demonstrate how the material being presented is related to future job skills or arouse curiosity. The instructional content would then use few or relatively simple visuals (compensate for spatial ability). Again, this module would be designed solely for low verbal reasoners.

The above instructional strategy is not conclusive. However, it is a researchable strategy that is meant to be (1) empirically derived and (2) based on a rationale to facilitate learning for a group of students who might be able to profit from some kind of alternative module.

During the alternative module validation phase, regression analyses would be run, against the criterion, on various combinations of main track and alternative modules which had been assigned to (1) all students, (2) predetermined sub-groups of students, and (3) random sub-groups of students. By comparing these regression slopes, and weighing the cost involved in alternative module development against the payoff derived, decisions could be made about whether to continue alternative module implementation or continue with main track materials only (see Figure 1, Points "F" and "G").

If the validation phase bears out the proposed hypothesis, then the alternative module would be assigned to this sub-group of students as the "primary" module at the start of the instructional unit. It is in this area of module assignment, module accounting, and record keeping that computer managed systems are most valuable.

Summary

The above procedure is an attempt to synthesize some of the recent thinking about ATI studies and alternative instructional module development. After all, ATI analyses and resulting alternatives are neither always appropriate nor cost effective and should be viewed as part of a larger picture. That is, they should be viewed as part of an empirically based decision model for alternative instructional module development.

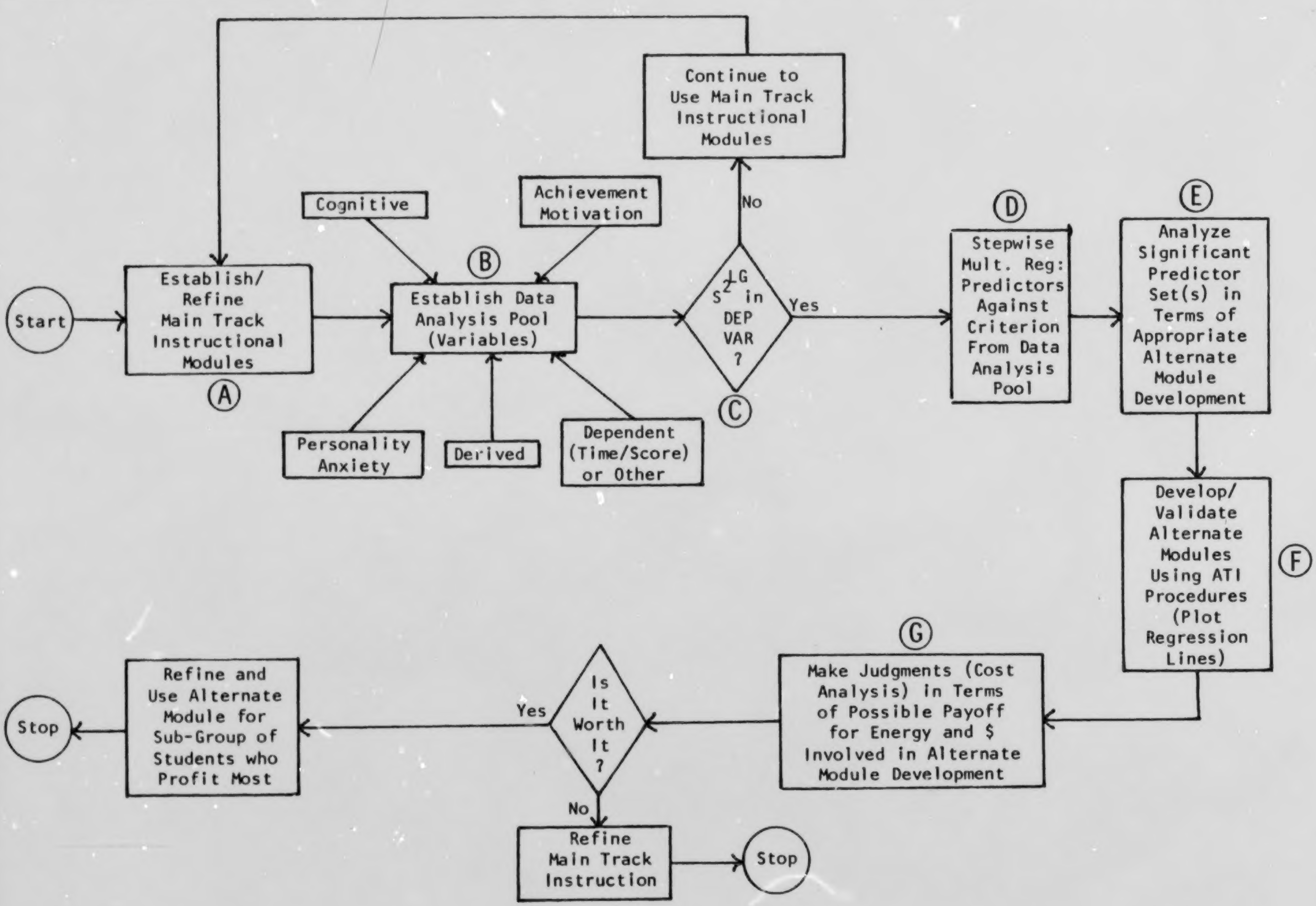

Start
Establish/ Refine Main Track Instructional Modules
Ⓐ
Cognitive
Achievement Motivation
Ⓑ
Establish Data Analysis Pool (Variables)
Personality Anxiety
Derived
Dependent (Time/Score) or Other
Continue to Use Main Track Instructional Modules
No
$S^{2}\perp G$ in DEP VAR ?
Yes
Ⓒ
Ⓓ
Stepwise Mult. Reg: Predictors Against Criterion From Data Analysis Pool
Ⓔ
Analyze Significant Predictor Set(s) in Terms of Appropriate Alternate Module Development
Develop/ Validate Alternate Modules Using ATI Procedures (Plot Regression Lines)
Ⓕ
Ⓖ
Make Judgments (Cost Analysis) in Terms of Possible Payoff for Energy and $ Involved in Alternate Module Development
Is It Worth It ?
Yes
No
Refine and Use Alternate Module for Sub-Group of Students who Profit Most
Stop
Refine Main Track Instruction
Stop

Figure 1 - A Data Based Procedure for Developing Alternative Instructional Modules

References

Cronbach, L. J. and Snow, R. E., *Aptitudes and Instructional Methods.* New York: Irvington Publishers, 1977.

deCHARMS, R. and Muir, M. S. Motivation and social approaches. *Annual Review of Psychology* 1978, *29*, 91-113.

Bracht, G. H. Experimental factors related to aptitude-treatment interactions. *Review of Educational Research*, 1970, *40*, 627-645.

Leherissey, B. L., O'Neil, H. F., Jr. and Hansen, D. H. Effects of memory support on state anxiety and performance in computer-assisted learning. *Journal of Educational Psychology*, 1972, *62*, 413-420.

Leherissey, B. L., O'Neil, H.F., Jr., and Hansen, D.H. Effect of Anxiety, response mode, and subject matter familiarity in computer assisted learning. *Journal of Educational Psychology*, 1973, *64*, 310-324.

Parkhurst, P.E. Generating meaningful hypotheses with aptitude-treatment interactions. *AV Communication Review*, 1975, 171-183.

Rhetts, J. E. Task, learner, and treatment variables. *Journal of Educational Psychology*, 1974, *66*, 339-347.

Snow,R. E. Individual differences and instructional theory. *Educational Researcher*, 1977, *6*, 11-15.

Snow,R. E. and Salomon, G. Aptitudes and instructional media. *AV Communication Review*, 1968, *16*, 341-357.

Spielberger, C.D., Conceptual and methodological issues in anxiety research. In C. D. Spielberger (Ed.), *Anxiety: Current Trends in Theory and Research.* New York: Academic Press, 1972.

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

IN SEARCH OF A BETTER WAY TO ORGANIZE INSTRUCTION: THE ELABORATION THEORY

Charles M. Reigeluth
Syracuse University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Michael Simonson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Working draft for an AECT presentation
March 8, 1979

ABSTRACT

It is becoming increasingly evident that there are many deficiencies in the hierarchical task analysis approach to designing instruction. The Elaboration Theory of Instruction was developed as an alternative that overcomes those deficiencies. The Elaboration Theory calls for beginning the instruction with a special kind of overview--one that epitomizes the instructional content rather than summarizing it. Then it calls for elaborating on that overview in a specific way==by adding detail or complexity in "layers" across the entire breadth of the content, one layer at a time, until the desired level of detail or complexity is reached.

In Search of a Better Way to Organize Instruction:
The Elaboration Theory

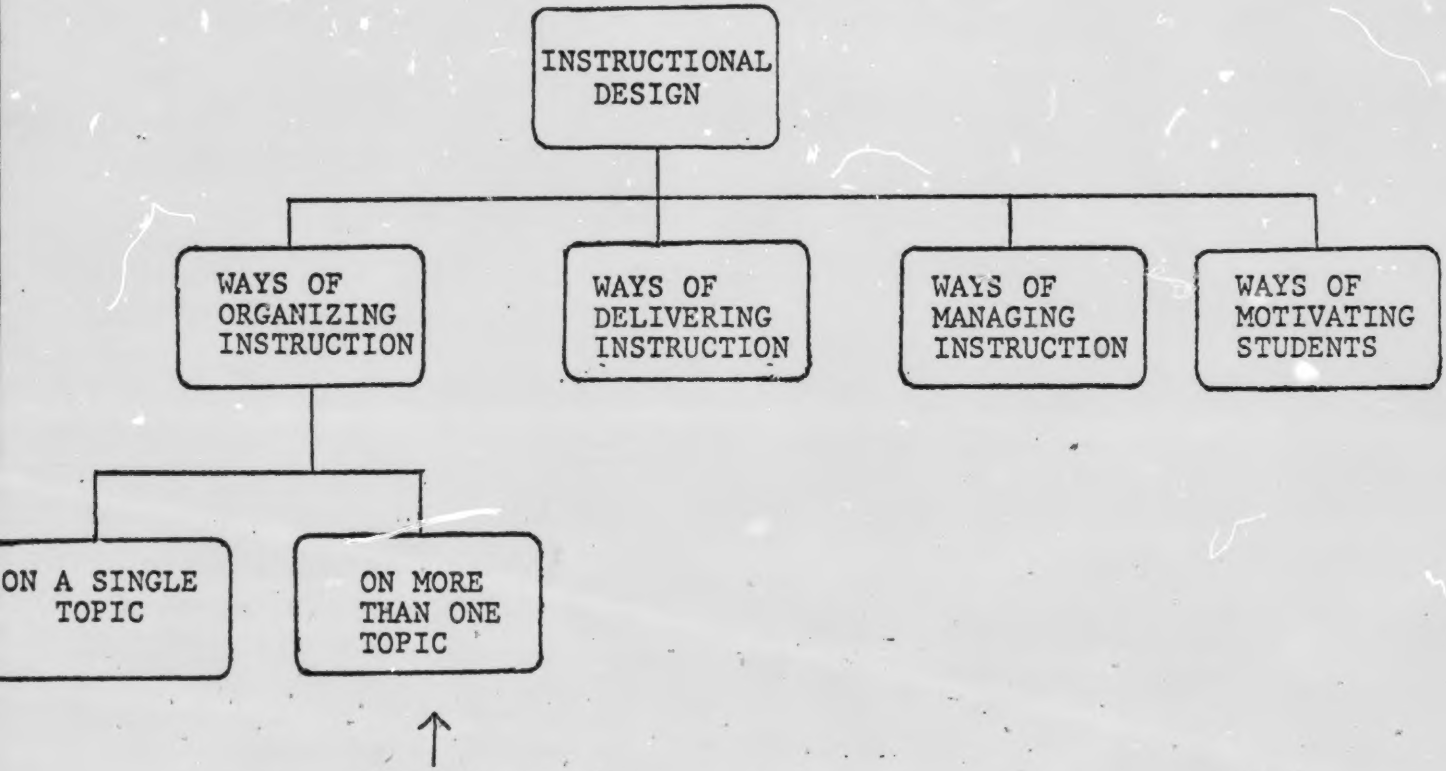
The Elaboration Theory of Instruction is an alternative to the standard way of organizing instruction based on a hierarchical task analysis. The hierarchical organization results in an instructional sequence that begins with highly fragmented, small pieces of the subject-matter content. Many educators have found the fragmentation to be demotivating. Many educational psychologists have found the parts-to-whole sequence to be inconsistent with much knowledge about how learning occurs most effectively--namely schema theory and its predecessor, subsumption theory. And many instructional designers have found that learning hierarchies represent a very incomplete basis upon which to make decisions about sequencing the instruction, mostly because learning hierarchies are only one aspect of the structure of subject-matter content. All this is not to deny that learning prerequisites exist nor to negate that they are important--they do exist and they are important. Rather this affirms that learning prerequisites are not a sufficient basis for organizing a whole course: our knowledge must progress beyond the hierarchy.

Context

Instructional design theory is concerned with methods of instruction. It is helpful to conceptualize two different levels of methods for organizing instruction: micro strategies, which are methods for organizing the instruction on a single topic (i.e., on a single concept, principle, etc.), such as generalities, examples, practice, and feedback; and macro strategies, which are methods for organizing those aspects of instruction which relate to more than one topic, such as sequencing (ordering), synthesizing (integrating), and summarizing (previewing and reviewing) all of the topics.

The Elaboration Theory of Instruction is a partial theory of instruction--it does not attempt to deal with all aspects of instruction. It does not deal with micro strategies for organizing instruction, although it can be and is being expanded to include such strategies (Reigeluth & Merrill, in preparation). For a good description of micro strategies, see Merrill's component display theory (Merrill, Reigeluth, & Faust, 1979; Merrill, Richards, Schmidt, & Wood, 1977). Also, the Elaboration Theory does not attempt to deal with strategies for delivering instruction (e.g., media selection), nor does it deal with strategies for managing instruction. Finally, most aspects of strategies for motivating students are not included within the present domain of the Elaboration Theory. But all of these aspects of instructional theory will be integrated with the Elaboration Theory in the foreseeable future. The Elaboration Theory of Instruction presently deals only with macro strategies for organizing instruction (see Figure 1).

Insert Figure 1 about here


INSTRUCTIONAL DESIGN
WAYS OF ORGANIZING INSTRUCTION
WAYS OF DELIVERING INSTRUCTION
WAYS OF MANAGING INSTRUCTION
WAYS OF MOTIVATING STUDENTS
ON A SINGLE TOPIC
ON MORE THAN ONE TOPIC

The Elaboration Theory is here

Figure 1. The context of the Elaboration Theory in relation to other aspects of instructional design theory.

An Analogy

A good introduction to the nature of the Elaboration Theory of Instruction is an analogy with a *zoom lens*. Taking a look at a subject matter "through" the Elaboration Theory approach to organizing instruction is similar in many respects to looking at a picture through a zoom lens on a movie camera.

A person starts with a *wide-angle view*, which allows one to see the major parts of the picture and the major relationships among those parts (e.g., the composition or balance of the picture), but without any detail.

The person then *zooms in* on a part of the picture. Assume that, instead of being continuous, the zoom operates in steps or discrete levels. Zooming in one level on a given part of the picture allows the person to see the major subparts of that part and the major relationships among those subparts. After having studied those subparts and their interrelationships, the person could then *zoom back out* to the wide-angle view to review the other parts of the whole picture and to review the context of this part within the whole picture.

The person continues this pattern of *zooming in* one level to see the major subparts of a part and *zooming back out* for context and review, until the whole picture has been seen at the first level of detail. Then the person can follow the same zoom-in/zoom-out pattern for the second level of detail, the third level, and so on, until the desired level of detail is reached.

In a similar way the Elaboration Theory of Instruction starts the student with an overview of the major parts of the subject matter, it elaborates on one of those parts to a certain level of detail (the first level of elaboration), it reviews the overview and shows the context of that part within the overview (an expanded overview), it continues this pattern of elaboration/expanded overview for each part of the overview until all parts have been elaborated one level, and it follows the same pattern for further levels of elaboration. Of course, it must be remembered that the zoom-lens analogy is just an analogy and therefore that it has non-analogous aspects. One such dissimilarity is that all the detail of the picture is actually present (although usually not noticed) in the wide-angle view, whereas the detail is not there at all in the overview of the subject matter.

The general-to-detailed organization prescribed by the Elaboration Theory helps to insure that the learner is always aware of the context and importance of the different topics that are being taught. It allows the learner to learn at the level of detail that is most appropriate and meaningful to him or her at any given state in the development of one's knowledge. And the learner never has to struggle through a series of learning prerequisites that are on too deep a level of detail to be interesting or meaningful at the initial stages of instruction. Very few learning prerequisites (if any) exist at the level of the overview. As a learner works one's way to deeper levels of detail, increasingly complex prerequisites will need to be introduced. But if they are only introduced at the level of detail at which they are necessary, then there will only be a few prerequisites at each level; and the learner will want to learn those prerequisites because he or she will see their importance for learning at the level of detail that now interests him or her.

Unfortunately, the zoom lens approach has not been used much in instruction, in spite of its fundamental simplicity and intuitive rationale. Many textbooks begin with the "lens" zoomed in to the level of detail deemed appropriate for the intended student population, and they proceed--with the "lens" locked on that level of detail--to pan across the entire subject matter. This has had unfortunate consequences for synthesis, retention, and motivation. Many instructional developers begin with the lens zoomed all the way in and proceed in a highly fragmented manner to pan across a small part and zoom out a bit on that part, pan across another small part and zoom out a bit on it, and so on until the whole scene has been covered and to some limited degree integrated. This has also had unfortunate consequences for synthesis, retention, and motivation. And some educators have intuitively groped for an elaboration-type approach with no guidelines on how to do it. This has resulted in a good deal less effectiveness than is possible for maximizing synthesis, retention, and motivation.

The major reason for the lack of utilization of the zoom lens approach in instruction is probably that the hierarchical approach was well-articulated and was a natural outgrowth of a strong behavioral orientation in educational psychology. This in effect put "blinders" on most of the few people who were working on instructional design strategies and methodology.

THE ELABORATION THEORY

The elaboration theory of instruction states that if cognitive instruction is organized in a certain specified way, then that instruction will result in higher levels of learning, synthesis, retention, and affect, all other things being equal. There is one limitation to this theory: the smaller the amount of interrelated subject-matter content, the less difference it will make. With a small enough number of topics, it doesn't make any difference how you sequence them, whether you synthesize them, or whether you summarize them (as long as there are no learning prerequisite relationships among them). The following is a description of that "certain specified way" of organizing instruction, which is called the Elaboration Model of Instruction.

The Elaboration Model of Instruction starts by presenting knowledge at a very general or simplified level--in the form of a special kind of overview. Then it proceeds to add *detail* or complexity in "layers" across the entire breadth of the content of the course (or curriculum), one layer at a time, until the desired level of detail or complexity is reached. It is important to emphasize that the Elaboration Model prescribes a special kind of overview, and it prescribes a special way in which the elaboration is to occur.

The Epitome

We do not like to use the word "overview" because its meaning is very vague--it means different things to different people. Also, we believe that a certain specific kind of overview is superior to other kinds. Among other things, our overview must *epitomize* the subject matter that is to be taught, rather than summarizing it. Hence, we have named it the epitome (ē·pĭt′ō·mē). An epitome has two "critical characteristics" which distinguish it from other types of overviews: (1) it *epitomizes* the subject matter of the course (or curriculum) rather than summarizing it, and (2) it has a single "*orientation*"--it emphasizes a single type of content.

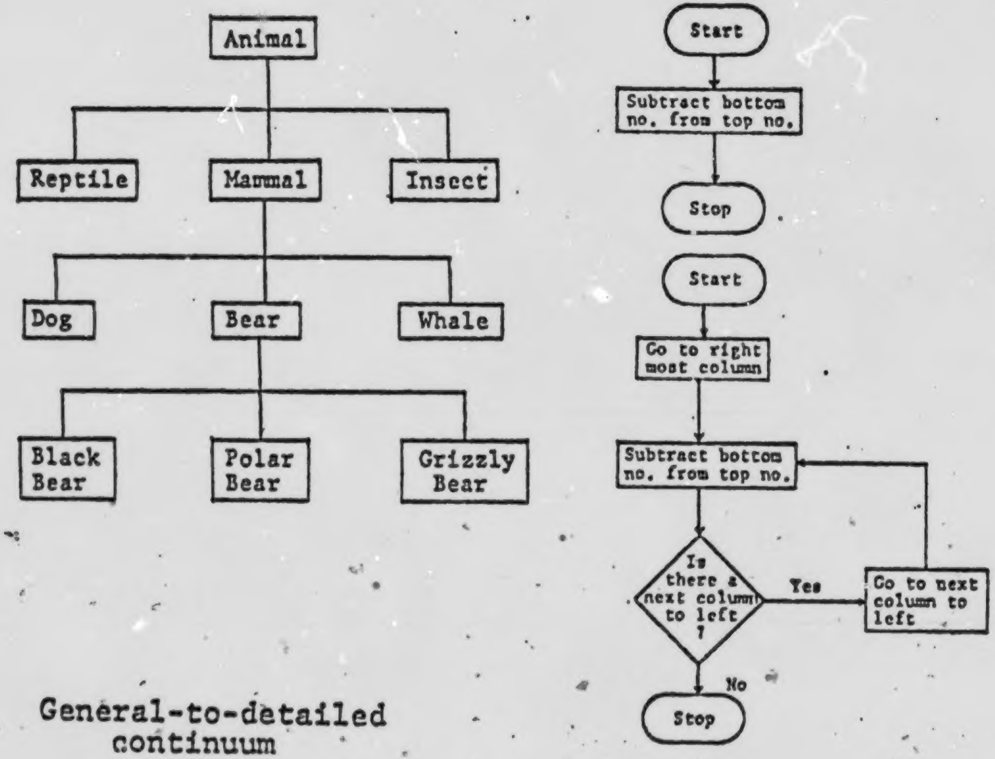
With respect to *epitomizing* the subject matter of the course (or curriculum), an epitome is formed by "boiling down" the course content to its essence. It does not preview all of the course content; rather it presents a few fundamental topics that convey the *essence* of the entire content. Those topics are chosen or derived in such a way that all the remaining course content provides more detail or more complex knowledge about the epitome. Although an epitome is very general, it is not purely abstract. Since "general" and "abstract" are often confused, this distinction will be discussed in greater detail shortly.

With respect to having an orientation, the epitome emphasizes any one of three types of content: concepts, procedures, or principles. A concept is a set of objects, events, or ideas that have certain characteristics in common. Knowing a concept entails being able to identify, recognize, classify, or describe what something is. A procedure is a set of actions that are intended to achieve an end. It is often referred to as a skill, a technique, or a method. Knowing a procedure entails knowing how to do something. A principle is a change relationship--it indicates the relationship between a change in one thing and a change in something else. It describes causes or effects by identifying what will happen as a result of a given change (the effect) or why something happens (the cause). These three different emphases are referred to as a conceptual orientation, a procedural orientation, and a theoretical orientation, respectively; and the orientation is selected on the basis of the general goals or purpose of the course (or curriculum). All three types of content may appear in the epitome, but one type receives primary emphasis; and the epitome is formed by epitomizing the orientation content and then introducing whatever of the other two types of content are highly relevant. More will be said about this below.

We mentioned above that an epitome is very general but is not purely abstract. The terms "general" and "abstract" are often confused. It is helpful to think of three continua: (1) general to detailed, (2) simple to complex, and (3) abstract to concrete. The first two are very similar to each other, but the third is very different (see Figure 2).

Insert Figure 2 about here

The general-to-detailed continuum refers primarily to a continuum formed by subdividing things (concepts or procedures) or by lumping things (concepts or procedures) together. "General" has breadth (things lumped together), while "detailed" is usually narrow (subdivisions). In Figure 2a "polar bear" is a more detailed concept than "animal." The simple-to-complex continuum refers primarily to a continuum formed by adding or removing things (principles or procedures). "Simple" has few things, while "complex" has many things. In figure 2b, the procedure for subtracting multi-digit numbers is more complex than the procedure for subtracting single-digit numbers. Additional complexity can be added by introducing sub-procedures for "borrowing" when the top number is smaller than the bottom number. The abstract-to-concrete continuum refers to tangibility, and there are two major types of tangibility. First, generalities are abstract, and instances are usually concrete--the definition of a tree is abstract, while a specific tree (an object) is con-


Animal
Reptile
Mammal
Insect
Dog
Bear
Whale
Black Bear
Polar Bear
Grizzly Bear
Start
Subtract bottom no. from top no.
Stop
Start
Go to right most column
Subtract bottom no. from top no.
Is there a next column to left ?
Yes
Go to next column to left
No
Stop

General-to-detailed continuum

(a)

Note: These are just two points on the continuum. For others, see Merrill, P.F. (1978).

Simple-to-complex continuum

(b)

Generalities

1. A rise in the supply of a good causes a drop in the price of the good in a free market.
2. A pen is an instrument that is used for writing with ink.
3. A German finite verb is placed at the end of a subordinate clause.

Instances

1. The record potato harvest this fall caused a drop in the price of potatos. +
2. (Teacher holds up a pen and says to class, "This object is a pen.")*
3. Er sagt, dab er das nicht tun will.

+Technically, this is a statement about the instance and is not the instance itself.

*Unlike No. 3, the instance is not actually present here on this page.

Abstract-to-concrete continuum

(c)

Figure 2. Illustrations of three continua that are often confused.

crete. This is the most important abstract-to-concrete continuum for instructional theory. Second, some concepts are considered abstract because their instances are not tangible. "Intelligence" is a good example of an abstract concept. This second abstract-to-concrete continuum is largely irrelevant for our purposes.

On the basis of these distinctions, an epitome is always either very general or very simple--it must be, in order to epitomize the instructional content. But it should never be purely abstract. According to Merrill's Component Display Theory, it should contain the following for each topic it presents: a generality (e.g., the definition of a concept), some instances of that generality (e.g., examples of the concept), and some practice for the student in applying the generality to new instances. An epitome usually contains about six (plus or minus three) topics--that is, about six different generalities, along with their instances and practice items. These topics may be any combination of concepts, procedures, and/or principles.

Figure 3 illustrates the nature of a theoretical epitome and of a conceptual epitome for an introductory course in ecomonics, and Figure 4 illustrates the nature of a procedural epitome for a course in literature.

Insert Figures 3 and 4 about here

----

1. The law (principle) of supply and demand.
    a. The principle of what causes changes to occur in the quantity demanded and the quantity supplied (price changes).
    b. The principle of why prices change in a free market economy.
2. The principle of why changes occur in supply schedules or demand schedules.
3. The concepts of supply, supply schedule, and supply curve.
4. The concepts of demand, demand schedule, and demand curve.
5. The concept of changes in quantity supplied or demanded.
6. The concept of changes in supply schedules or demand schedules.
7. The concept of equilibrium price.

Practically all principles of economics can be viewed as elaborations on the law of supply and demand, including those that relate to monopoly, regulation, price fixing, and planned economies.

Figure 3a. The instructional content for a theoretical epitome for an introductory course in economics.

---

1. Definition of economics
2. Definitions of subdivisions of economics:
    a. Definition of macro economics
    b. Definition of micro economics
    c. Definition of comparative economics
    d. Definition of international economics
    e. Definition of labor economics
    f. Definition of managerial economics.

Practically all concepts in economics can be viewed as elaborations on these concepts (i.e., as further subdivisions--either parts or kinds--of these concepts).

Figure 3b. The instructional content for a conceptual epitome for an introductory course in economics.

1. There are three major stages in the multidimensional analysis and interpretation of creative literature:
    a. Identifying the elements of the dramatic framework--character and plot.
    b. Combining the elements into composites appropriate for analysis of their literal meaning--analysis of character in terms of plot.
    c. Figuratively interpreting the elements--symbolism through character, mood, tone.
    d. Making a judgment of worth--personal relevance, universality.

(This procedure is simplified by introducing only two elements for the analyses in a and b, three in c, and two in d. It is further simplified by introducing only those procedures and concepts necessary for the analysis and interpretation of a short poem. Complexity is later added by increasing the number of elements used in each stage of analysis or interpretation and by introducing procedures and concepts needed for analyzing and interpreting more complicated types of creative literature.)

2. Concepts necessary for performing the procedure in 1.
    a. character
    b. plot
    c. symbolism
    d. mood
    e. tone
    f. universality

Figure 4. The instructional content for a procedural epitome for an introductory course in literature.

A Level-1 Elaboration

A level-1 elaboration is a part of the instruction that provides some more detailed or complex knowledge on a topic (or set of topics) that was introduced to the student in the epitome. It should not include all of the more detailed or complex knowledge on that topic. Rather, a level-1 elaboration should be an epitome of all of the more detailed or complex knowledge on that topic, just as zooming in one level provides a slightly more detailed wide-angle view of one part of the whole picture. There may be as many level-1 elaborations as there are topics in the epitome, but there does not have to be a one-to-one correspondence. It is possible that each of the topics in a level-1 elaboration may elaborate to some extent on all of the topics in the epitome or perhaps even on a relationship among those topics.

The depth to which a level-1 elaboration should elaborate on a part of the epitome is somewhat variable (i.e., the discrete levels on the zoom lens are variable, not always constant and equal in the amount of detail added). The most important factor for deciding on the depth of a given level-1 elaboration is student learning load. It is important that the student learning load be neither too large nor too small, for either will impede the instruction's efficiency, effectiveness (especially for retention), and appeal. The number of topics that represent the optimal student learning load will vary with such factors as student ability, the complexity of the subject-matter topics, and student prefamiliarity with the topics. The breadth of a level-1 elaboration will usually be fairly difficult to adjust. Hence optimizing the student learning load in a given elaboration can often be done only by varying the depth of that elaboration.

Figure 5 illustrates the nature of a level-1 elaboration on the theoretical epitome in Figure 3, and Figure 6 illustrates the nature of a level-1 elaboration on the procedural epitome in Figure 4.

Insert Figures 5 and 6 about here

Other Elaborations

A level-2 elaboration is identical to a level-1 elaboration except that it elaborates on a topic (or set of topics) introduced in a level-1 elaboration rather than in the epitome. In a similar manner, a level-3 elaboration provides more detail or complexity on a topic (or set ot topics) introduced in a level-2 elaboration, and so on for elaborations at deeper levels of detail/complexity. In all cases, an elaboration at one level of detail/complexity should be an epitome for all the elaborations that elaborate on it.

According to this kind of organization, elaborations that are on the same level are very different from each other

1. Principle of increasing marginal costs as an explanation for the shape of the supply curve.
2. Principle of profit maximization for individual firms.
3. Procedure of marginal analysis to arrive at profit maximization.
4. Concepts of fixed and variable costs.
5. Concepts of total, average, and marginal costs.
6. Concepts of break-even point and shut-down point.

Figure 5. The instructional content for a level-1 elaboration on the theoretical epitome in Figure 3a. It elaborates on the supply aspect of the law of supply and demand by presenting more complex principles that relate to supply.

1. How to identify other elements of the dramatic framework--setting, perspective, and language.
2. How to combine the elements into composites appropriate for analysis of their literal meaning--(1) analysis of character, plot, and setting, (2) analysis of persepctive, character, and plot, and (3) analysis of language.
3. Concepts of setting, perspective, and language.
4. Concepts of types and patterns of imagery (in language).
5. Procedure for analyzing imagery.
6. Concept of prosody.
7. Procedure for analyzing prosody.

Figure 6. The instructional content for a level-1 elaboration on the procedural epitome in Figure 4. It elaborates on stages a and b (which must be elaborated at the same time because of their interrelatedness) by adding elements that need to be identified (in stage a) and analyzed in combination (in stage b).

with respect to the instructional content they contain (i.e., their topics are very different from each other); but elaborations that are on different levels are very similar to each other with respect to their instructional content (i.e., their topics are very similar) because each level has the same content as the previous levels, only at a level of greater detail/complexity.

Expanded Epitome

After each elaboration, the instruction presents a summarizer and an expanded epitome, equivalent to the zoom-out-for-context-and-review activity in the zoom lens analogy. The summarizer is comprised of a concise generality for each topic presented in the elaboration. The expanded epitome (a) synthesizes the topics presented within the elaboration (internal synthesis) and (b) shows the relationship of those topics (and relationships) to the rest of the topics (and relationships) that have been taught (external synthesis).

Summary of the Elaboration Model

In summary, the Elaboration Model is as follows (see Figure 7). First, the epitome is presented to the student. Then a level-1 elaboration is presented to provide more detail on an aspect of the orientation content in the epitome (that aspect which is most important or contributes most to an understanding of the whole orientation structure). Next a summarizer and an expanded epitome are presented. Another level-1 elaboration and its summarizer and expanded epitome are presented. This pattern of level-1 elaboration followed by its summarizer and expanded epitome continues until all aspects of the orientation content that were presented in the epitome have been elaborated one level. Then a level-2 elaboration is presented to provide more detail on an aspect of the orientation content that was presented in one of the level-1 elaborations. As always, this elaboration is followed by a summarizer and an expanded epitome. This pattern continues until all of the aspects of the orientation content presented in all of the level-1 elaborations have been elaborated one level (unless the objectives of the course or the nature of the subject matter exempt a level-1 elaboration from being further elaborated). Additional levels of elaboration are provided in the same manner--an elaboration followed by a summarizer and an expanded epitome--until the level of detail/complexity specified by the objectives is attained in all aspects of the orientation structure (and supporting structures) of the course.

Insert Figure 7 about here

USING THE ELABORATION MODEL

We have developed a fairly detailed set of procedures for designing instruction according to the Elaboration Model (Reigeluth, et al, 1978). A major part of those procedures is analyzing the instructional content as to four different types

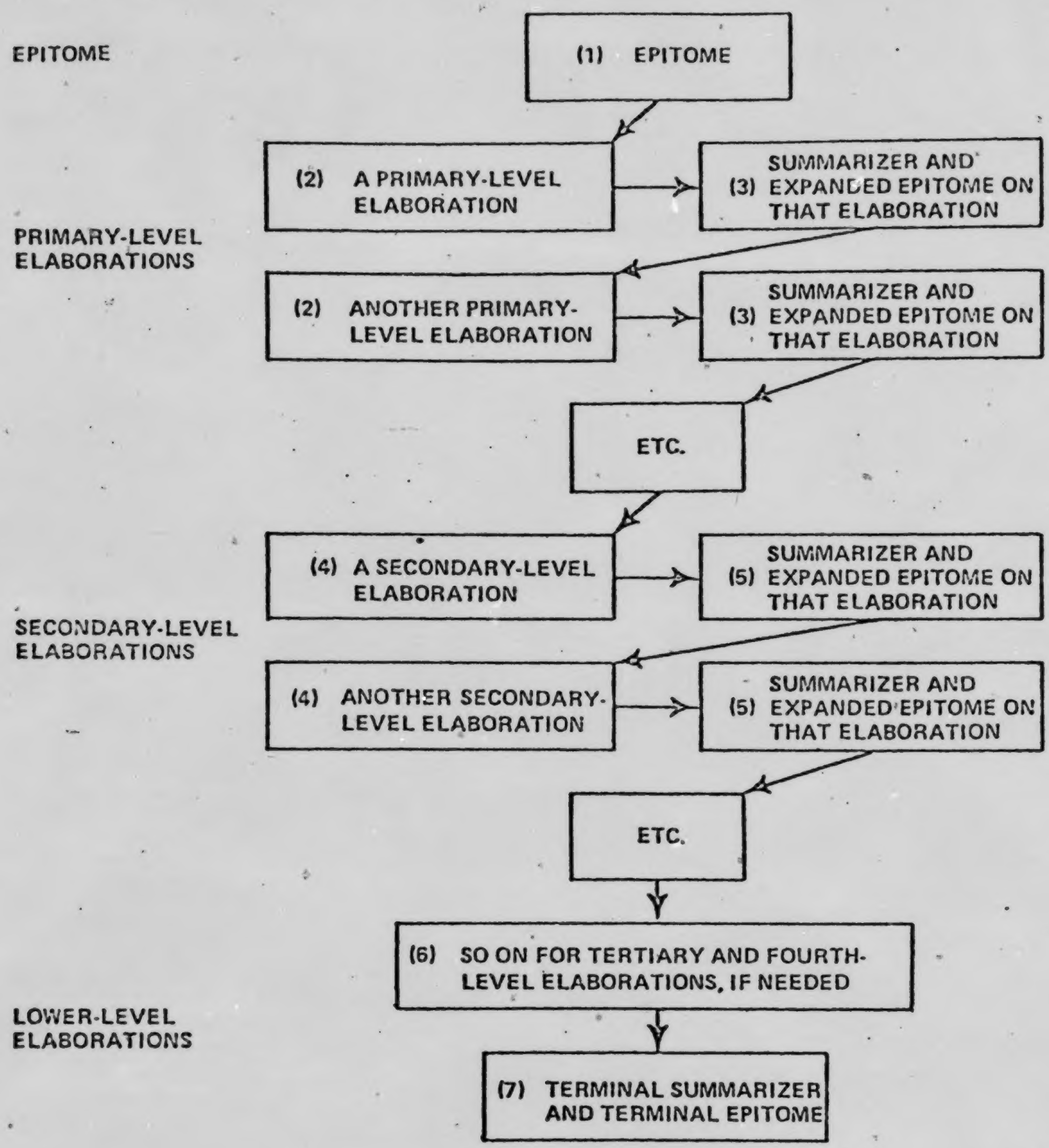

EPITOME
(1) EPITOME
PRIMARY-LEVEL ELABORATIONS
(2) A PRIMARY-LEVEL ELABORATION
(3) SUMMARIZER AND EXPANDED EPITOME ON THAT ELABORATION
(2) ANOTHER PRIMARY-LEVEL ELABORATION
(3) SUMMARIZER AND EXPANDED EPITOME ON THAT ELABORATION
ETC.
SECONDARY-LEVEL ELABORATIONS
(4) A SECONDARY-LEVEL ELABORATION
(5) SUMMARIZER AND EXPANDED EPITOME ON THAT ELABORATION
(4) ANOTHER SECONDARY-LEVEL ELABORATION
(5) SUMMARIZER AND EXPANDED EPITOME ON THAT ELABORATION
ETC.
(6) SO ON FOR TERTIARY AND FOURTH-LEVEL ELABORATIONS, IF NEEDED
LOWER-LEVEL ELABORATIONS
(7) TERMINAL SUMMARIZER AND TERMINAL EPITOME

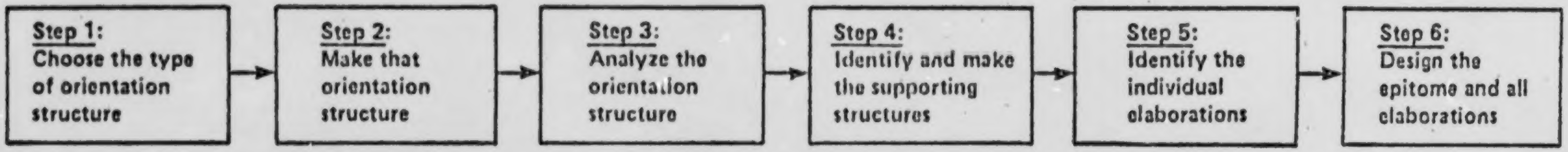
Figure 7. A diagramatic representation of the Elaboration Model of Instruction.

of subject-matter structures[1]: conceptual, procedural, theoretical, and learning. (Learning structures show learning prerequisite relations within the subject matter.) It is beyond the scope of this paper to describe and illustrate each of these four types of structures. The interested reader is referred to Reigeluth, Merrill, & Bunderson, 1978.

There are six major steps for designing instruction according to the Elaboration Model (see Figure 8). *First*, one must select an orientation--either conceptual, procedural, or theoretical--on the basis of the goals or purpose of the instruction. *Second*, one must develop an orientation structure for that orientation. It depicts the orientation content (either concepts, procedures, or principles) in the most detailed/complex version that the student needs to learn. This is a form of content analysis or task description. *Then* the orientation structure is analyzed in a systematic manner to determine which aspect(s) of the orientation content will be presented in the epitome and which aspects will be presented in each level of elaboration. In this way the "skeleton" of the instruction is developed on the basis of epitomizing and elaborating a single type of content.

Insert Figure 8 about here

The *fourth* major step is to embellish the "skeleton" by


Step 1: Choose the type of orientation structure
Step 2: Make that orientation structure
Step 3: Analyze the orientation structure
Step 4: Identify and make the supporting structures
Step 5: Identify the individual elaborations
Step 6: Design the epitome and all elaborations

Conceptual
Procedural
Theoretical

2a. Conceptual
Develop all useful parts and kinds taxonomies, select the most important one(s), and combine into a matrix (if possible).

2b. Procedural
Identify all useful steps and alternative paths to be learned, and combine them into a procedural structure.

2c. Theoretical
Identify all important principles to be learned, and combine them into a theoretical structure.

Decide on the aspects of the orientation structure that should comprise the epitome and each level of elaboration.

3a. Conceptual
Prune the orientation structure to form the epitome. Add back levels to form each level of elaboration.

3b. Procedural
Lump steps and branches together to form the epitome. Break apart each general step to form each level of elaboration.

3c. Theoretical
Prune the orientation structure to form the epitome. Use rank-order of importance and/or a parallel conceptual structure to identify principles for each level of elaboration.

Identify all unattained contextual, procedural, and explanatory goals, and make the corresponding supporting structures. Then identify all unattained learning prerequisites, and make their corresponding learning structures.

Divide each level of elaboration into its individual elaborations, according to the part of the epitome, or of an elaboration, that is being elaborated.

Design the epitome.
- general synthesizer
- constructs
- summarizer
- epitome (synthesizer)

Design each elaboration in the primary level
- general synthesizer
- constructs
- summarizer
- expanded epitome

etc.

Design the terminal epitome.
- orientation structure
- important supporting structures

Figure 8. The six-step design procedure for structuring the instruction in any course entailing cognitive subject matter.

adding the other two types of content at the lowest appropriate levels of detail. This is usually done by "nesting" the remaining subject-matter structures within different parts of the skeleton. (This may include some isolated structures or topics in the orientation content that did not fit in with the orientation structure in forming the skeleton.) Learning prerequisites are one of the considerations that enter in at this point.

Having allocated all of the instructional content to the different levels of elaboration, it is now important to establish the scope and depth of each individual elaboration that will comprise each level. The scope is usually predetermined by the orientation topic and its necessary supporting topics, although two orientation topics can be lumped together into a single elaboration, and it would be possible (though not advisable) to add extra supporting topics. The depth is then determined on the basis of achieving an optimal student learning load, as described above.

Sixth and finally, some of the internal structure of each elaboration within each level can be planned. The sequence of topics within an elaboration is decided on the basis on contribution to an understanding of the whole orientation structure (but within the constraints of learning prerequisites), and the locations of synthesizers and summarizers are also determined.

This concludes the "macro" design process, at which point the "micro" design process begins--decisions as to how to organize the instruction on a single topic.

THE NEED FOR RESEARCH

The model and procedures as described above have undergone very limited field-testing and virtually no research. It may turn out that having a complete expanded epitome after every single elaboration is inefficient and unnecessary (especially after lower-level elaborations). It may also turn out that it is unnecessary for a student to study all level-1 elaborations before proceeding to a level-2 elaboration. This would have important implications for learner-controlled selection and sequencing of topics--a student could now truly follow one's interests in approaching a subject matter. This would be particularly valuable in adult and continuing education contexts.

It is also likely that a large, full-scale field test of the design procedures will reveal more effective and efficient ways to design instruction according to the model.

The Elaboration Model as developed to date is a tentative move in a much-needed direction. It does not yet have the maturity and validation of the currently used approaches to instructional design, but the need for alternatives should be clear. And there is great potential for the Elaboration Model to meet that need.

SUMMARY

It is becoming increasingly evident that there are many deficiencies in the hierarchical task analysis approach to organizing instruction on the macro level. The resulting instructional designs are usually fragmented, demotivating, inconsistent with learning theory, and at best a very incomplete basis for organizing instruction. The Elaboration Theory of Instruction was developed as an alternative that overcomes these deficiencies. But it is emphasized that it is but a partial theory of instruction--it only deals with macro strategies for organizing instruction (see Figure 1 above).

The zoom-lens analogy was presented as an introduction to the nature of the Elaboration Theory of Instruction. A person starts with a *wide-angle view* and then proceeds to *zoom in* one level for detail on a part of the picture and *zoom out* for review and context. After the whole first level has been studied, the same zoom-in/zoom-out pattern is followed for the second level of detail, and so on until the whole picture has been studied at the desired level of detail. Alternative approaches to designing instruction were contrasted in terms of this analogy, and their deficiencies were mentioned.

Next, the Elaboration Model of Instruction was described. It starts by presenting a special kind of overview, called the "*epitome*", which (1) epitomizes the instructional content rather than summarizing it and (2) has a single "orientation" (a single type of content). The single orientation may be either conceptual, procedural, or theoretical. And to epitomize the content, the epitome contains one or two very general/simple topics from the orientation content; but the epitome should not be purely abstract--the epitome should be formed in such a way that concrete instances and practice can be presented on each topic. Some of the other two types of content may also be included in the epitome if they are highly relevant to the orientation content topic(s).

After presenting the epitome, the model calls for adding detail or complexity in "layers" across the entire breadth of the content, one layer at a time, until the desired level of detail/complexity is reached. First, level-1 elaborations are presented, each of which provides more detail/complexity on one aspect of the epitome. After all aspects of the epitome have been elaborated in this manner, level-2 elaborations are presented. Each of them elaborates on one aspect of a level-1 elaboration. Additional levels of elaborations are added as necessary to reach the level of detail/complexity called for by the objectives. The amount of content in each elaboration must be carefully planned so as to represent an optimal student learning load. Also, each elaboration, regardless of level, is followed by a summarizer and an expanded epitome.

Next the procedure for using the elaboration model in the design of new instruction was summarized. The six major steps are shown in Figure 8 above.

Finally, the need for continued research, field-testing, and development of the Elaboration Model was emphasized. The Elaboration Model as developed to date is a tentative move in a much-needed direction. The need for alternatives to the hierarchical approach should be clear, and there is great potential for the Elaboration Model to meet that need.

REFERENCES

Merrill, M.D., Reigeluth, C.M., & Faust, G.W. The Instructional Quality Profile: A Curriculum Evaluation and Design Tool. In H.F. O'Neil, Jr., (Ed.) *Procedures for Instructional Systems Development*, New York: Academic Press, 1979.

Merrill, M.D., Richards, R.A., Schmidt, R.V., & Wood, N.D. *The Instructional Strategy Diagnostic Profile Training Manual*. San Diego: Courseware, Inc., 1977.

Reigeluth, C.M., & Merrill, M.D. *Instructional Theory and Models: A Road to Better Instruction.* In preparation.

Reigeluth, C.M., Merrill, M.D., & Bunderson, C.V. The Structure of Subject-Matter Content and Its Instructional Design Implications. *Instructional Science*, 1978, *7*, 107-126.

Reigeluth, C.M., Merrill, M.D., Wilson, B.G., & Spiller, R.T. *Final Report on the Structural Strategy Diagnostic Profile Project*. A final report submitted to the Navy Personnel Research and Development Center, San Diego, California, July 1978.

FOOTNOTE

[1] A subject matter structure is something which shows a single kind of relationship that exists within a subject matter. Figure 2a shows part of a subject-matter structure.

A Consumer's Guide to Common Flaws in Research[*]

Dennis M. Roberts

The Pennsylvania State University

[*] The present paper is a revised and condensed version of a paper "Common Flaws In Research Design" by William Rabinowitz and Dennis Roberts published in the NABTE Journal, 1977, #4.

Planning the Research Study

The purpose of research is to shed light on important questions phrased in testable forms. In general, one seeks to establish functional or causal relationships between independent and dependent variables. However, designing studies to establish these relationships or causes involves many potential hazards. In any case though, the investigator has an obligation to plan the study in order to maximize the interpretability of the findings. Unfortunately, there are many basic flaws that can seriously hinder results interpretation. The present brief paper is a sequential walk through a simple research setting-- a two group experiment-- to indicate common (but avoidable) design difficulties that will jeopardize the usefulness of the obtained data.

Common Flaws

Figure 1 presents a flow chart type diagram of several important steps in the research process. Obviously, it is kept simple-- many details of actual design execution have been omitted because of space limitations. Assume, for purposes of the remaining discussion, that "the" important question is "Will the use of electronic calculators improve elementary school students' attitudes towards mathematics?"

Insert Figure 1 about here

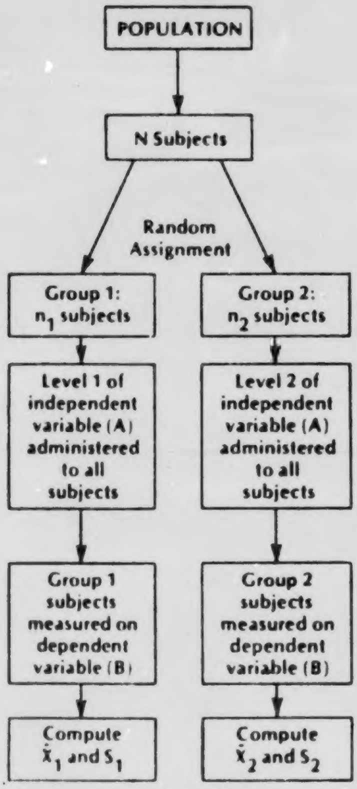
---

1. Failure to select subjects from the population to which generalizations are to be made.

This item deals specifically with what Campbell and Stanley (1963) refer to as external validity. Clearly, the question implies that an answer is

Figure 1. Steps and possible design flaws in an experiment to study the effect of treatment variable A on variable B.

FLOW CHART


POPULATION
N Subjects
Random Assignment
Group 1: n_1 subjects
Group 2: n_2 subjects
Level 1 of independent variable (A) administered to all subjects
Level 2 of independent variable (A) administered to all subjects
Group 1 subjects measured on dependent variable (B)
Group 2 subjects measured on dependent variable (B)
Compute $\bar{X}_1$ and S_1
Compute $\bar{X}_2$ and S_2

STEPS	FLAWS
The relevant population is identified	Failure to select subjects from the population to which investigator expects to generalize.
N Subjects are selected from the population.	Failure to use randomization in assigning subjects to groups.
Subjects are divided randomly into two groups.	Failure to use an adequate number of subjects.
Subjects in both groups receive identical treatment in all respects *except* that different levels of the independent variable (A) are administered to each group.	Failure to maintain control over manipulation of the independent variable (A).
Subjects in both groups are measured on the dependent variable (B). Measurements are made in the same way for all subjects	Failure to use reliable and/or valid measures of the dependent variable (B).
Statistics are computed for each sample.	Failure to use an appropriate statistical test.
Hypothesis test conducted to decide whether the difference between $\bar{X}_1$ and $\bar{X}_2$ is due to sampling variability or to the effects of variable A	

$H_0: \mu_1 - \mu_2 = 0$
$H_1: \mu_1 - \mu_2 \neq 0$

desired for the overall population of elementary school students. Therefore, there should be evidence in the methods section that shows students at varying elementary grade levels are included in the sample. Secondly, there must be evidence that the students from these several grade levels were drawn (for the study) in a way to assure a representative sample. If the sample differs in important ways from the population, the results may not generalize to the population of interest as stated in the research question.

2. Failure to use randomization in assigning students to experimental and control groups.

It is absolutely essential that both groups be equivalent at the onset of the experiment. Differences between the groups on the dependent variable that appear at the conclusion of the study will invariably be interpreted as resulting from the impact of the independent variable. However, this conclusion cannot be justified if groups were different initially in important (and unknown) ways. Unfortunately, there is no way to absolutely guarantee equivalence of both groups initially. However, random assignment does ensure equivalence within the limits of chance. Since random fluctuations (chance aberrations) are estimatable, random assignment offers a way to control the initial makeup of the two groups. If the method section does not give specific indications that students were assigned at random to the two groups, the entire set of results are very suspect.

3. Failure to use an adequate number of students.

If the hope is to support the notion that use of calculators will improve attitudes towards mathematics, then we certainly want to reject the null hypothesis (μ Attitude with Calculators = μ Attitude without Calculators) if our notion is actually correct. In this situation, rejecting the null in the study when our position is the true state of affairs is a correct decision to have reached. This type of correct decision is called the power of a statistical test. Unfortunately, unless your independent variable is *very* potent, power is greatly influenced by sample size. Also, unfortunately, the ability to build a certain degree of power into a study-- while doable-- is fairly complicated. A paper by Hopkins (1974) may be helpful to the interested reader. In the present study, it is doubtful that calculators are going to work miracles on students' sttitudes towards mathematics. Thus, a failure to use a "sufficient" sized sample is certainly doomed to a retention of the null hypothesis.

4. *Failure to maintain control over manipulation of the independent variable.*

A necessary assumption is that students in the two groups were, in fact, treated differently *throughout* the period of the study. In our context, this *requires* that no control students have access to calculators during the experiment. While this may be controllable during the math classes, control students might have access to and use calculators in other classes (science?) and/or outside of school. Such contamination is very likely to *reduce* the differences between the two groups on the dependent variable. What will look like an ineffective *independent* variable actually disguises the possibility that the calculator did have an effect; however, inadvertently in *both* groups.

5. *Failure to use reliable and/or valid measures of the dependent variable.*

The original question posed stated that the important dependent variable of interest is "attitudes toward mathematics". Assuming for a moment that such a construct exists, then at issue is whether the construct can be measured consistently. Without some data being presented to demonstrate reliability of the measure in *this* particular research setting, then it is unlikely that differences between the groups on the dependent measure will occur. Assuming reasonable reliability, however, a more important issue is whether the dependent measure in the study is, in fact, measuring attitudes toward mathematics. Obviously, any comparison between the groups on a measure that isn't tapping the important construct of interest is useless and certainly no test of the hypothesis we might have generated from our research question (i.e., calculators improve attitudes towards mathematics).

6. Failure to use an appropriate statistical test.

One area where flaws shouldn't occur but do quite often is the inappropriate use of certain statistical tests. The emphasis here is not on using a test that is less powerful than others but rather on using a test incorrectly. A common example here is using analysis of covariance (ANCOVA) to make statistical adjustments for the fact that groups were not equivalent initially. What this means (in all probability) is that a statistical technique is supposed to correct for non-randomization of students to the experimental and control groups. Unfortunately, ANCOVA can't make non-random assignment random; no statistical test can correct for that difficulty. ANCOVA has two sets of assumptions that need to be met in order to be interpreted properly-- there are those for analysis of variance (ANOVA) and regression. If random assignment has been made and there is a reasonable correlation between a covariate (measured--incidentally-- before the experiment begins) and the dependent variable, then ANCOVA may reduce the error term, hence, giving a better chance to reject the null hypothesis. The interested reader is referred to an article on ANCOVA by Kennedy (1977) for elaborations on this topic.

Closing Comments

The present focus has been on several common flaws that occur in even the simplest research design. Obviously, many other flaws may creep into research studies such that the output data are rendered suspect. First, many studies do not use an appropriate research design to handle the hypotheses being tested. Second, many research questions are formulated in such a way that no data will be able to support the hypotheses. And third, many investigations are conducted without proper consideration of the relevant variables that might have an influence on the dependent measures. All these additional problems plus the others briefly outlined above pose threats to both the internal and external validity of research investigations. Careful design and control, plus a good measure of common sense, are the best ways in empirical studies to insure meaningful data from one's research efforts.

References

Campbell, D., and Stanley, J. Experimental and quasi-experimental designs for research. Handbook of Research on Teaching, N. L. Gage, editor. Skokie, Ill.: Rand McNally, 1963.

Hopkins, K. D. Preventing the number-one misinterpretation of behavioral research, or how to increase statistical power. Journal of Special Education, Volume 8, 1974.

Kennedy, J. J. The use and misuse of analysis of covariance. NABTE Review, 1977.

title • REANALYSIS OF RESEARCH STUDIES IN *AVCR*

authors • DENNIS E. SHERIFF
Assistant Professor
Faculty of Instructional Technology
Northern Illinois University

• JOHN A. HORTIN
Doctoral Student
Instructional Technology Programs
Northern Illinois University

correspondence • Dr. Dennis E. Sheriff
Faculty of Instructional Technology
Northern Illinois University
DeKalb, Illinois 60115

- Reanalysis of Research Studies in *AVCR*
- PURPOSE

A desideratum of behavioral science research is to provide technically correct and statistically accurate answers to all research questions investigated within an experiment. Focusing upon this criterion, researchers in the area of instructional technology have identified many valid research questions, designed experiments that isolate variables to be analyzed, selected statistical models to study quantifiable data, and drawn inferences to larger populations which were under investigation. However, using empirical observation methodology, it was revealed that 108 studies which appeared in the first twenty-five volumes of *AV Communication Review* selected a univariate analysis of variance model to test experimental hypotheses. Thus, the primary purpose of study was twofold: (1) to compute the interrelationships between independent and criterion variables for the 108 studies previously identified; and, (2) to report these calculations in the form percents to the profession.

- RATIONALE

A tendency exhibited by most researchers who published research findings in the journal cited hereinbefore was to test experimental hypotheses based upon a statistical model designed to identify significant differences. Very few studies were found that reported practical significances or significant relationships. Clark and Snow (1975, p. 392) suggested that alternative experimental designs should be employed to investigate additional research questions, i.e., practical significances and significant relationships. While manipulating the design to control variables is desirable, the robustness of interrelationships between dependent and independent variables is paramount to advancing the knowledge base within the field of instructional technology.

• METHOD

Reanalyzing a typical analysis of variance study, the researcher attempts to statistically suggest that there are no differences between criterion variables, while totally disregarding the statistical association between the independent and criterion variables. Both statistical and practical significance tests can be computed from the data reported in most F-test tables. Hays (1963, p. 325) suggested that a statistical index called ω^2 (Greek omega, squared) can be employed to estimate the *proportion of variance in Y accounted for by X.* Omega, squared values are not difficult to interpret. They are somewhat similar to correlation coefficients in that each is a mathematical expression of the relationship that exists between two or more variables. These values are expressed in the form of a continuum ranging from 0 to 1.00. (Omega, squared can never assume a negative value.) When ω^2 is 1.00, the independent variable is congruent to the criterion variable. However, when ω^2 assumes a zero value, the knowledge of the independent variable does not in any way reduce the uncertainty about the criterion variable.

In most studies that were reanalyzed, Hays' (1963, p. 327) formula for computing ω^2 had to be abandoned because most researchers reported incomplete F-test tables. Thus, McNamara's (1978, p. 51) alternate version was applied to the 108 experimental studies that appeared in the first twenty-five volumes of *AVCR*. His equivalent formula was reported as

$$\text{est. } \omega^2 = \frac{(k-1)\cdot F-(k-1)}{k-1\cdot F+(N-k)+1}$$

where F is the numerical value for the ANOVA test, N is the total number of individuals in the sample, and k is the number of levels of the independent variable.

RESULTS

The following results were germane to the study.

1. Seven studies were not included in this reanalysis because they failed to report information necessary to calculate ω^2. In all seven cases, the researchers failed to report the degrees of freedom.
2. Two of the 108 studies reported ω^2 values.
3. Forty-eight percent of the 108 studies appeared in Volumes 19-25.
4. A total of 1,019 F-tests were computed in the 108 identified studies.
5. About 23 percent of all F-tests reanalyzed produced ω^2 values between .0001 - .0099.
6. Sixty-eight percent of all ω^2 values fell within the .0001 - .0499 incurement.
7. Only 1.5 percent of the F-tests produced ω^2 values larger than .7000.

REFERENCES

Clark, R.E. and Snow, R.E. Alternative designs for instructional technology research. *AV Communication Review*, 1975, *23*, 373-394.

Hays, W.L. *Statistics*. New York: Holt, Rinehart, and Winston, 1963.

McNamara, J.F. Practical significance and statistical models. *Educational Administration Quarterly*, 1978, *14*, 48-63.

Torkelson, G.M. AVCR - One quarter century evolution of theory and research. *AV Communication Review*, 1977, *25*, 317-358.

Research for the Practitioner

Brenda H. White
University of Pittsburgh

Prepared for the AECT Conference, March, 1979.

RESEARCH FOR THE PRACTITIONER

A common response when research is mentioned is "so what"! We all have read research studies which address seemingly meaningless questions or which disappoint with their inconclusive or inaccurate findings. This can easily lead one to be mistrustful of all research. To exercise caution in reviewing research and in accepting the conclusions drawn is a healthy attitude. It is one that insists that the questions asked are valid and that the conclusions drawn are significant. It is one which resists fads and seeks truth. It is this attitude that I hope you will maintain during this presentation. However, to distrust and discount all research on the basis of some bad experiences would be akin to throwing the baby out with the bathwater.

What, then, can research do for you? Why bother to come to a presentation on research for practitioners? I suspect that one of the reasons you've come to this session - and to this Conference - is that you are interested in new ideas and want to keep up to date with the field.

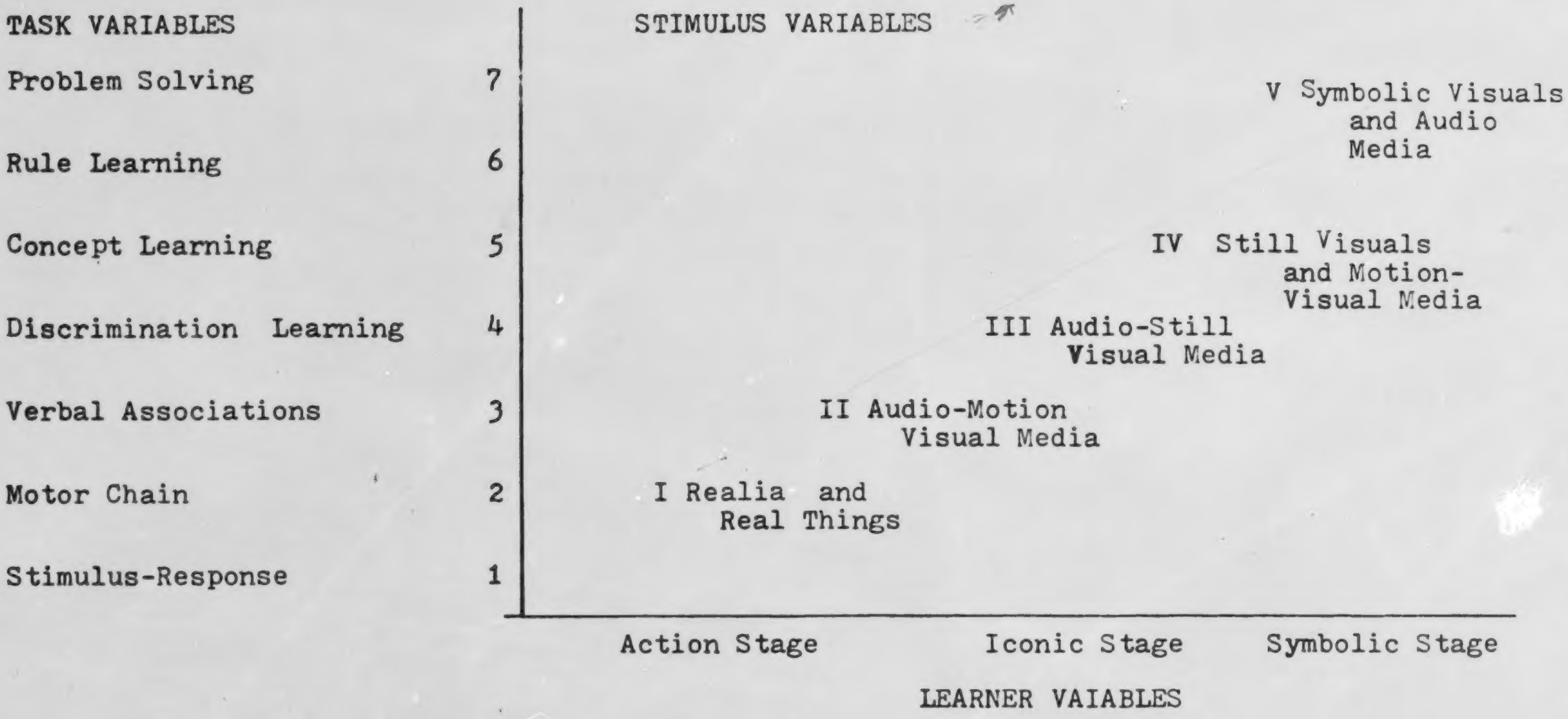
Another reason for an interest in research may be to overcome the isolation in which many of us work. Media specialists in the public schools usually practice alone, one to a school and have few contacts with others who understand the work. Contacts with fellow specialists are usually limited to a few in-service days a year and regional meetings or conferences such as this one. Knowing something about what is being done by researchers can help us feel part of the larger whole.

It is also useful to be able to quote research findings when proposing program changes to administrators. Nancy Polette, in In-Service, advised using research findings to sell them on the value of conducting in-service programs. They have been trained to be skeptical, to demand proof of ideas and research findings can provide such evidence.

You can use research to sell ideas in other ways as well. For the last few years accountability has been a by-word in educational circles. With the continuing shrinkage in enrollment (or stabilization in the public schools at lower levels than the peak years of the sixties), value for dollars expended is a significant concern. A study such as Janet Simone's (A Feasibility Study for the Centralized Processing of Print and Non-Print Materials for the Elementary School Libraries in the Allegheny Intermediate Unit) could help illustrate such accountability for an administrator.

The Allegheny Intermediate Unit consists of 46 public school districts in the county of Allegheny (in Pennsylvania). The study was limited to this area and is not generalizable to a different geographical area, although the method can become a decision making process for more general application.

This study utilized the survey research method with data collected through questionnaires and telephone interviews. The data collected was processed and analyzed by computer using the Statistical Package for the Social Sciences. Cost analysis showed that centralized processing, at the time of this study, would have saved from $6.52 to $6.96 per item


TASK VARIABLES
STIMULUS VARIABLES
Problem Solving 7
Rule Learning 6
Concept Learning 5
Discrimination Learning 4
Verbal Associations 3
Motor Chain 2
Stimulus-Response 1
V Symbolic Visuals and Audio Media
IV Still Visuals and Motion-Visual Media
III Audio-Still Visual Media
II Audio-Motion Visual Media
I Realia and Real Things
Action Stage
Iconic Stage
Symbolic Stage
LEARNER VAIABLES

Graphic illustration of the D-Model for selection

Figure One

over the cost of individual district processing.

Another area in which research can improve the return for dollars and for time invested is that of selection of media. When inappropriate media is selected, both money and time are lost. Money that should have been spent on materials more suited to the task has been misspent. Time that a student spends using inappropriate materials is time wasted. When one considers all that a child must learn in his/her school years, there is little room for a haphazard selection of materials. To improve the ability of practitioners to select appropriate media for described tasks, Barbara Dompa developed the D-Model. (See Figure One.) This model is designed to facilitate the interaction of task, stimulus, and learner components in the selection of media in order that the messages necessary to permit the development of the learner's competencies will be discernible.

As part of this research an experiment was designed to test whether or not there would be a significant difference between novice librarians using the D-Model and novice librarians using selection aids in the choices made. The experimental design was as follows:

	Pretreatment Phase	Treatment Phase	Post-Treatment Phase
Control group	C_1	S	C_2
Experimental group	E_1	M	E_2

The pretreatment phase was a pretest administered to both groups. The treatment phase for the control group was a one-hour review of selection aids. For the experimental group it was a one-hour explanation of the D-Model. Directly following the treatment, both groups received a post-test in which they were asked to identify appropriate media to meet eight behavioral objectives.

Analysis of the data generated revealed a statistically singificant difference between the performances of the control and experimental groups, with the experimental group out-performing the control group. From this it appears that an ability to use the D-Model can lead to improved selection of media for specific tasks, and, thus, to improved effectiveness.

Yet another use of research is in suggesting methods for evaluating and improving media services. In today's world of tight budgets, it is necessary that media specialists and librarians be aware of the effectiveness of their programs. One way to test program effectiveness and to enhance awareness of media services is to conduct a self-evaluation study.

In 1973 David Leortscher studied media center services in Indiana. For this study he prepared and administered a questionnaire regarding services offered by the school media center. The questionnaire was administered to teachers and media center staffs in nine typical Indiana schools. The questions covered the areas of accessibility, awareness, production, instruction and consultation. Media center staffs were asked to respond, on a scale of 0 to 3, whether or not a service was provided and how frequently it was provided.

Teachers were asked to respond as to whether they received the service and to what degree they were satisfied with the service. In general, media center staffs agreed among themselves about the services they were offering, but the teachers did not agree. This research showed that in the best of circumstances, only about 40 to 60% of the teachers agreed with the media staff about the services offered. (Another interesting finding was that individual differences among teachers affected library use as much as did the subject taught.)

This research indicated that a priority task for media specialists is to ensure that teachers and other users know what services are, in fact, offered. One way to alert them is to use a survey like Loertscher's, since answering a survey can enhance teacher awareness as can a follow-up report. As an outgrowth of his study Loertscher has developed, with Janet Stroud, a service called "PSES -- Purdue Self-Evaluation System for School Media Centers". From a catalog the specialist selects the survey items to use in the school. For a fee, the user will be supplied with a reproducible questionnaire and machine readable answer sheets for each person included in the survey. The answer sheets are analyzed by computer at Purdue and the results are reported to the school.

In another study, JoAnn Rogers studied the effect of teachers knowledge of media skills upon their use of media centers. She developed the School Media Center Fundamentals Test in order to determine whether "a relationship exists between possession of necessary skills, attitude toward media formats, and frequency of use of media center resources, print and non-print." Her study should have an impact on the provision and content of both pre-service training and in-service training for teachers. The media specialist can increase teacher use of media center materials by providing appropriate in-service training.

An important group of users that should not be neglected as one strives to increase the visibility of the media program is students. Jacqueline Mancall has studied the resources used by high school students in preparing independent study projects. Library use was determined by studying bibliographies of student papers, by a questionnaire surveying the students who prepared the papers, by a questionnaire surveying the teachers involved, and by interviews with the librarians. Mancall discovered that students prefer to use monograph materials with little consideration for the currency of the materials, and that materials other then monographs and journals must be specifically brought to students' attention. An encouraging aspect of the study was that these students saw libraries as places where they can and do receive assistance.

Instruction in the use of library/media centers is an area that is becoming more popular for research. Some time ago, Corlett at the University of Portland, measured library skills as a predictor of academic success. She administered the Feagley, Curtiss, Gaver and Greene Library Orientation Test and the Brown-Holtzman Test to a sample of 81 college students enrolled in freshman English classes in a small metropolitan university (fall, 1969-70 school year). The criterion of academic success was the grade point average attained at the end of the semester. She performed correlation analysis and multiple regression analysis. Her conclusions, from the statistical analysis of the data, were that the Library Orientation Test appeared valid for forecasting success in college. This study should be replicated to extend the population to which the findings can be generalized.

Also in the area of instruction is a study of instructional methods, by Clark Chie-Yuen Wong. He studied the effectiveness of two methods of orienting college students to the Learning Materials Center of the Community College of Denver. He used five groups with a total of 162 subjects and six librarians. The criterion measures were a written quiz and a performance exercise, both of which had been pilot tested. He used the Randomized Equivalent Groups Posttest Only true experimental design:

	ASSIGNMENT AFTER ORIENTATION	ASSIGNMENT BEFORE ORIENTATION
LECTURE METHOD	G_1	G_3
SLIDE-TAPE METHOD	G_2	G_4

Groups 1 and 3 received orientation by lecture; groups 2 and 4, by slide-tape method. Groups 1 and 2 received the performance exercise assignment after the orientation; groups 3 and 4, before the orientation.

Major findings were:

1. Students receiving orientation achieved significantly higher on the quiz than those not receiving orientation.
2. Neither method of orientation resulted in significantly different achievement on the quiz.

3. Students receiving the assignment before the orientation achieved significantly higher than those receiving the assignment after the orientation.

It should be noted that this study can only be generalized to community college students.

In 1978, Helen Gothberg reported on a study of alternative methods of teaching basic reference to library school students. The study compared the lecture method to an audio-tutorial method of instruction. Audio-tutorial units included a recorded tape, a bibliography of required readings, suggested topics for discussion, additional information sheets, a study guide and a worksheet. The results of the study showed no significant differences in the understanding of theoretical concerns, regardless of the method of instruction; no differences in the knowledge of the basic reference tools, and no differences in student satisfaction with the methodology. However, audio-tutorial students reported more satisfaction with their own performance and knowledge than did the lecture method students. A constraint on the study was contamination through an exchange of information between the control and experimental groups. Students in the lecture group believed that the audio-tutorial students had superior notes.

In a review of research, William H. Allen analyzed trends in instructional media research as reported in AV Communication Review and the Review of Educational Research during the five-year periods of 1963-73 and 1973-77. Based on his review of the published material concerning research, he concluded that three areas receive most attention from researchers: nature and characteristics of media factors related to the design and development of media, and the relationship of media and media design to learning. He reported a move from evaluative research to research concerning the nature of media, particularly that of the pictorial image.

Commonly, aptitude-treatment interaction studies have attempted to isolate learning variables and the effect of aspects of media upon these variables. Ann DeVaney Becker has written a provocative article concerning such methods of investigation of instructional media and its effect on learning. She comments that the stimulus-response model used in such studies introduces distortion through the experimenter's attempt to control the intervening variables. According to her, the S-R model neglects the effects of intention, action and situation as variables in the learning process. She asserts that learning may be better described through language than through numbers and the reduction of data to statistics. In place of the S-R model, Becker suggests that field research techniques be used, such as case studies utilizing interviews, record or document surveys, direct observation, indirect observation and participant observation in order to provide more relevant data. The reliability of these studies may be checked by observing the agreement between different observers at different time, the agreement between different observers at the same time and the agreement between observations made by the same observer at different times.

I hope that I have presented a persuasive case for the value of research in helping us to find answers to the problems that face us day by day. If so, I want to make this final point: research is not conducted in a vaccuum. Researchers need to hear from the field. Practitioners should provide feedback concerning what research is needed and concerning the appropriateness of the research which has been done. Only then can the contribution of research be most appreciated and utilized.

BIBLIOGRAPHY

William H. Allen. TRENDS IN INSTRUCTIONAL MEDIA RESEARCH. University of Southern California.

Ann DeVaney Becker. "Alternative Methodologies for Instructional Media Research." AV COMMUNICATION REVIEW, 25(Summer, 1977), 181-194.

Donna Corlett. "Library Skills, Study Habits and Attitudes, and Sex as Related to Academic Achievement." EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 34 (1974), 967-969.

Barbara F. Dompa. A SYSTEM FOR SELECTING MEDIA FOR SCHOOL LIBRARY COLLECTIONS BASED ON THE MATCHING OF TASK, STIMULUS AND LEARNER VARIABLES. Dissertation. Case Western Reserve University, 1975.

Helen M. Gothberg. "A Study of the Audio-Tutorial Approach to Teaching Basic Reference." JOURNAL OF EDUCATION FOR LIBRARIANSHIP, (Winter, 1978), 193-202.

David V. Loertscher. MEDIA CENTER SERVICES TO TEACHERS IN INDIANA SENIOR HIGH SCHOOLS, 1972-1973. Dissertation. Indiana University, 1973.

David V. Loertscher and Janet Stroud. PSES: PURDUE SELF-EVALUATION FOR SCHOOL MEDIA CENTERS. Hi Willow Research and Publishing, 1976.

Jacqueline C. Mancall. RESOURCES USED BY HIGH SCHOOL STUDENTS IN PREPARING INDEPENDENT STUDY PROJECTS: A BIBLIOMETRIC APPRAOCH. Drexel University.

Nancy Polette. IN-SERVICE: SCHOOL LIBRARY/MEDIA WORKSHOPS AND CONFERENCES. Metuchen, NJ, Scarecrow Press, 1973.

JoAnn V. Rogers. TEACHERS AND MEDIA RESOURCES IN SELECTED APPALACHIAN SECONDARY SCHOOLS. Dissertation. University of Pittsburgh, 1976.

Mary Janet Simone. A FEASIBILITY STUDY FOR THE CENTRALIZED PROCESSING OF PRINT AND NON-PRINT MATERIALS FOR THE ELEMENTARY SCHOOL LIBRARIES IN THE ALLEGHENY INTERMEDIATE UNIT. Dissertation. University of Pittsburgh, 1977.

Clark Chie-Yuen Wong. COMPARATIVE EFFECTIVENESS OF THE LECTURE AND SLIDE-TAPE APPROACH FOR ORIENTATION IN THE USE OF LEARNING MATERIALS CENTER. Dissertation. University of Colorado, 1975.

PART II:

LEARNER CHARACTERISTICS AND PRESENTATION ORGANIZATION RESEARCH

Eye Movement Research and the Interaction Between Television and Child-related Characteristics

Lois J. Baron

Assistant Professor of Education

Concordia University

1455 de Maisonneuve Boulevard West

Montreal

Quebec

H3G 1M8

Canada

(514) 879-7351 - Office

(514) 739-0458 - Home

Synopsis

The relationship of eye movements while watching television, reading ability, cognitive style, and mode of presentation was studied. Eighty-five third grade children were classified as good or poor readers (Gates-MacGinitie Reading Test). Cognitive style was measured by the Children's Embedded Figures which classified subjects as field dependent or independent. A Polymetric Eye Movement recorder was used to measure attention as children viewed seven static and seven dynamic segments from The Electric Company. The results indicated that the effects of these factors varied from segment to segment although several overall effects achieved significance. It was concluded that eye movement research is a useful tool for assessing individual differences across instructional modes.

Eye Movement Research and the Interaction Between Television and Child-related Characteristics

Lois J. Baron

Research (Gagne, 1970; Salomon and Snow, 1968; Snow and Salomon, 1968; Salomon, 1972) has stressed the importance of the three-way interaction between the medium, the individual learner, and the learning task. This study exemplifies what Salomon and Clark (1977) have called a "fusion" of what had previously been examined through linear investigations of main effects. The research performed here can be described in terms of the definition of media research as the investigation of the psychological and instructional effects of media on the responses of individuals. This particular study describes, through the analysis of eye movements measured during the viewing of both dynamic and static television presentations, the perceptual analysis of individuals possessing differing degrees of field-independence-dependence and reading proficiency.

In perceiving, the individual constructs his world according to both the features of the stimuli before him, and the organization of his own cognitive structure. The gaze, as it is called

(Mackworth and Bruner, 1970; Mackworth and Morandi, 1967), is initially controlled by the display, but eventually becomes an instrument of thought. Studying the eye movements of field independents and field dependent individuals, and good versus poor readers lends insight into comparisons, if any, of developmental shifts from global processing to highly differentiated modes of perception.

Wolf (1971) and others (Fleming, 1969 and 1970) have supported the view that "the study of eye movements allows for an unusual opportunity for determining the reaction of viewers to stimuli materials." (Wolf, p. 113).

The research carried out here was based on the underlying assumption that a child's eye fixations reflect his cognitive approach to the task and that eye movements are indicative of attending behavior. Support for such an assumption is well founded in the literature (Mackworth and Bruner, 1970; Mackworth, 1967; Guba, Wolf, de Groot, Knemeyer, Van Atta, and Light, 1965; Gould, 1967, 1973; Gould and Schaffer, 1965a, 1965b; Gould and Dill, 1969; O'Bryan, 1969; O'Bryan and Boersma, 1971; Noton and Stark, 1971; Vurpillot, 1968). The research performed here investigated both individual difference and stimulus-related characteristics while at the same time examined the interaction between these factors. The primary objective of this research was to study by means of eye movement photography the attentional processes related to the organismic variables of field-independence-dependence and differing levels of reading proficiency, and to test whether there was an interaction between these factors and the element of movement in the stimulus presentation.

Eye movement research into the process of reading has generally found that reading proficiency varies with both the ability of the reader and the complexity of reading material (Tinker, 1947; Taylor, 1957, 1960; Conant, 1965; Fleming, 1969; Mackworth and Bruner, 1970; Nodine and Lang, 1971; Edelfat, 1975; Rayner, 1975). There is no doubt that eye movements can provide a way of observing the relative effectiveness with which a person reads (Taylor, 1960). Without attending to the stimuli it would be very difficult to process any information which in turn benefits reading operations (Mackworth, 1972). By providing descriptive information, this eye movement research attempted to lend insight into possibly different response availabilities between poor readers, bound to simple decoding, and good readers, analyzing at higher levels up the hierarchy of word and letter recognition.

There has not been much eye movement research as it relates to the area of field-independence-dependence although two particular studies (Conklin, Muir, and Boersma, 1968; and Boersma, Muir, Wilton, and Barham, 1969) have found differences in such scanning strategies as track length and information search. Originally defined by Witkin (1954, 1962, 1964, 1977), field-independence-dependence has been used to describe individual differences in overcoming embedding or potentially distracting contexts. Field independent persons possess a more analytic approach to tasks, in contrast to the global field approach characteristic of field dependent individuals (Witkin, 1962). The eye movement indices chosen for this study served as a description of those behaviors which are normally attributed to field independent individuals while at the same time specified the global characteristics of field dependent persons. An intention of this study was to assess the possibly embedding nature of various televised stimulus segments and to examine through eye movement analysis how embeddedness (or distractibility) may have affected the attention of field independent and field dependent children. Of particular importance to this study was the ability of the television code variable of motion to separate figure from ground. A number of arguments for action on the screen are found in research (Fowles, 1973; Rovet, 1974; Allen, 1975).

The appeal for research whose major aim is to examine stimulus characteristics as well as person-related variables had motivated this study.

Previous research comparing the various media as methods of presenting visualized instruction had been inconsistent in its conclusions, although it was generally acknowledged that no significant differences had resulted in most studies comparing the effectiveness of two or more visual media (Dwyer, 1973; Jamison, Suppes, & Wells, 1974). Salomon (1974) suggested that researchers must look at those variables peculiar to one medium itself. The importance of looking at the "symbol system unique to each medium" (p. 386), and how it interacts with certain learner characteristics results in more insights concerning individualization and the nature of the medium. Looking at the various forms of information representation and how they are operated on by an individual is more beneficial than studying differences across the technologies themselves. Studies of this sort touch the heart of the issue - "the relationship between the way information is externally represented and how it is internally processed" (Conway, 1970, p. 159).

Television possesses pictorial, symbolic, and verbal characteristics (Allen, 1970). According to some theorists (Olson, 1974; Olson & Bruner, 1974), the individual learner must become aware of the code within a symbol system in order to benefit from instructions. When producing a symbol system code for a particular learner it is important to take into account the possibility that too many irrelevant cues may also hinder learning. Rust (in Fowles, 1973) reiterated the point that such high appeal characteristics as rhythm, rhyme, and electronic bridges may be distracting as well as attention-getting.

In looking at the medium of television, one would classify movement as a code variable. Allen (1970) proposed the need for studies comparing action and non-action within the medium. He also stressed the necessity of including individual aptitude variables in research of this sort and advocated further investigation into those variables dealing with the perceptual processes and how they interact differentially toward certain media elements.

The appeal for research whose major aim was to examine stimulus characteristics as well as person-related variables motivated this study.

In the nature of formative research, existing segments of the Children's Television Workshop program The Electric Company were examined by means of eye movement photography. By doing so insights into the nature of the attentional behaviors of good and poor readers, and field independent, field dependent children were observed in interaction with televised segments from the program. As well, the nature of the stimulus element of movement and its resultant effect on attention was observed.

The independent factors of reading proficiency, field-independence-dependence, and dynamic versus static stimuli were examined in terms of their contribution to five eye movement measures - (1) orientation time to target or the length of time taken by a subject before making a fixation on a target area (word or sentences) once it appears on the screen (ORIENT) (2) percentage of fixations on target (FIXATION) (3) percentage of time on target (TIME) (4) left-to-right movements or directional attack (L→R) and (5) average length of fixation or fixation duration on target (AVERAGE). These measures have been used in past eye movement research studies which, in being selected for this research were considered useful indicators of how persons would interact with a stimulus presentation on the television screen, (particularly one involving "reading"). These indices provide information as to whether and how an individual is directing his eye fixations to targets on the screen, and also furnish insight into the effects of stimulus variables on eye movements.

In light of the literature concerning field-independence-dependence, reading proficiency, mode of instructional presentation, and eye movements, the following objectives or research questions were formulated:

1) By means of eye movement data, to gain more insight into the relationship between field-independence-dependence and reading proficiency.

2) By means of eye movement research, to determine the effects of the media element of movement on the visual patterning of good and poor readers and field independent and field dependent individuals.

3) By means of eye movement research, to further investigate differences in perception between good and poor readers.

4) Utilizing eye movement research, to examine perceptual differences between field independent and field dependent persons.

The general objective of this research study was the following: In the line of formative or ongoing field research using eye movements as a dependent variable, the research was performed in order to acquire more knowledge about the communication process between the medium and the individual.

It was hypothesized that good readers, field independent individuals, and dynamic stimuli would yield the more proficient and analytical scanning strategies exemplified by significantly faster orientation times to print, a significantly greater percentage of fixations on the target words, a significantly larger percentage of time on target, a significantly larger percentage of directional attack on target, and significantly shorter durations of fixation.

Method

Subjects

The subjects (Ss) were 85 third-grade pupils (42 boys, 43 girls) from eight elementary schools in the Scarborough, Ontario, Board of Education. The Ss were chosen according to their comprehension scores on the Gates-MacGinitie Reading Test, Primary C, Form 2 (1964). The test scores were obtained by examining the individual record file for each grade three student in the normal stream for each school.

Those students scoring at least one and one-half standard deviations above or below the Borough mean reading score were respectively assigned to good reader and poor reader groups. The Ss consisted of 41 good readers (15 boys, 26 girls) with a mean reading score of 40.82 and a mean grade equivalent score of 5.63, and 44 poor readers (27 boys, 17 girls) with a mean reading score of 10.93 and a mean grade equivalent score of 1.76.

Apparatus

A Polymetrics Eye Movement Recorder (Model V-1164-1) was used to record eye movement patterns (EMPs) in this study. Mackworth (1967) provided a detailed description of this line-of-sight recording equipment. Utilizing the corneal reflection method characteristic of this apparatus, a spot of light superimposed over the stimulus field was produced. These eye spots were recorded by a video camera as the Ss viewed the stimulus field on a nine-inch television screen housed at eye level approximately two feet from the subject. The location of eye spots was transmitted as digital signals to a PDP-9 computer.

Stimulus Materials

Segments from the Children's Television Workshop program The Electric Company (TEC) were used as a stimulus material. It was necessary to produce a stationary presentation from the already existing dynamic segments in order to assess possible differences in the quality of eye movements between dynamic and static modes of presentation. Doing so involved "freezing" the moving segments at particular intervals, juxtaposing the frozen bits (or segments) in the same order as that of the dynamic presentation, and dubbing the identical voice track over the resultant visuals. This method insured maximum content comparability between the two presentations.

A dynamic and a static presentation were created consisting of seven "bits" or segments each. In the nature of formative research, the segments were chosen for their individual stimulus qualities (e.g. computer or electronic bridges, animation, distracting versus non-distracting field).

Procedure

Administering the CEFT

The Children's Embedded Figures Test (CEFT) was administered individually to all Ss to assess cognitive style. A standardized procedure for administering the test was outlined in the Manual (Witkin, Ottman, Raskin & Karp, 1971). For the purposes of the analyses carried out in this study, a mean split was performed to divide those Ss who were considered field independent (FI) from those who were field dependent (FD).

Eye movement recording procedure.

Subjects entered the eye movement laboratory individually. The eye movement recording device was explained to the Ss. They were told that the E was interested in monitoring their television-watching and that by allowing the E to do so, they were contributing to the betterment of television programs for all children.

The next task was to assure that each S was in an appropriate and comfortable position for viewing the television set. The height of the chair was raised such that the S's forehead would rest comfortably on a headrest attached to the equipment. A bite-bar was utilized to minimize any head movements of the Ss. Each S also had an elasticized sweatband placed around his/her head for the same purpose.

Once in a proper position for watching television, calibration of the recording equipment took place. To calibrate, Ss were asked to look at four corner dots and a center dot affixed to the blank television screen. The recording equipment was then appropriately adjusted.

Each subject viewed the two stimulus presentations--dynamic and static. The order of presentation for each S was previously determined with the aid of a table of random numbers. Between presentations each S was given a two-minute rest.

Scoring procedure.

As outlined previously, five eye movement measures were chosen for investigating the questions and hypotheses of this research study.

Before measuring the eye movements, it was necessary to specify the criterion for defining an eye spot as a "fixation" for the purposes of this study. Taylor (1960) points out that the average length of fixation for a third-grader is .28 seconds (or 280 milliseconds) per eye spot. Owing to the fact that subjects in this experiment included children reading at levels as low as grade one, it was decided to establish .10 seconds or 100 milliseconds as the minimum amount of time necessary for an eye spot to be considered a fixation.

Before analyzing the data, it was necessary to insure validity of the measurements by eliminating defective recordings. Upon completing the data gathering, the video tapes of each subject's EMPs were reviewed without knowledge of the characteristics of that particular subject. Those segments lacking in calibration, as indicated by the eye spot being off the center spot of the screen following the termination of a segment, were excluded from analysis. This elimination process led to a possible situation in which for one particular subject there may have been only one or two good segments from which data could be analyzed. The program used in the final analyses comprised the necessary measures for the handling of missing data.

Preparation of Data for Analysis

In order to assess the eye movement patterns, information as to the exact time and location of each target was provided as data to the PDP-9 computer. The targets within each stimulus segment were a word or group of words, and the target field in which the eye movement measurements were made consisted of a four-sided figure surrounding these words. Targetting the dynamic segments consisted of either targetting certain areas during the presentation where a target letter or word would lie at a particular point in time (e.g. Try, Dry, Shy) or targetting a larger area in which whole words or sentences would appear (e.g. the Princess and the Frog). Targetting the static segments consisted of a similar process in which the unmoving words or letters were targetted for their position on the screen. Targetting constituted transmitting a pulse from each corner of a target to be changed to numerical coordinate information in the computer (in the same manner as an S's eye spot). Once this information became computer data, the location and analysis of subject data, already stored on computer tape, was carried out on the eye spots occurring within the time of target appearances. With the time and coordinate information available within the computer, a series of programs reduced the data in terms of the five eye movement indices of interest to this study--orientation times (in thirtieths of a second), fixations on target (as a percentage), time on target (as a percentage), average duration of fixation (in thirtieths of a second), and left-to-right (L→R) movement (as a percentage). The eye movement information was identified and categorized for each segment and subject. In the final analysis only data from the stimulus first presented to a subject (dynamic or static) was utilized in the calculations.

Statistical Treatment of Data

A regression procedure was used specifying the analysis of variance model. The particular program utilized was taken from the Statistical Analysis System (Service, 1972).

The program applied the method of least squares in fitting a linear model to the data. The variation attributable to each of the independent variables in the model (treated as classification variables) was examined as were the interactions of these variables on the five dependent or eye movement measures. Treating the variables as independent of each other was of concern to this study. For this reason, the partial sum of squares was the source of variation examined. The data analyzed was that available from each subject classified under either the dynamic or static condition, good reader or poor reader category, and field independent or field dependent classification.

Results

Regression analyses were calculated for each segment (referred to as ALL) as well as an overall analysis averaged over all segments together (MODE-AVERAGE). To account for the possibility that differences or lack of differences in the dependent measures could be due to long exposure to a stimulus, a regression analysis on information gathered during the early seconds of a segment's exposure was executed (INIT).

1) Orientation Time

Mode-Average data gathered throughout the segments revealed the only significant differences for this measure. Good readers did orient significantly faster than did poor readers (F= 5.07;

$p < .05$). Significant differences were found between field independent and field dependent subjects ($F = 4.74$; $p < .05$) and between males and females ($F = 5.49$; $p < .05$). Field independent subjects oriented to the target words significantly faster. Males also possessed significantly faster overall orientation times although in one particular cartoon segment girls oriented significantly faster to the target words ($F = 4.92$; $p < .05$).

Orientation time measures yielded three-way interaction and only on averaging data over all segments. The interaction of CEFT x SEX varied significantly with the presentation mode ($F = 4.63$; $p < .05$). The dynamic mode yield the shortest orientation time for both field dependent boys and girls.

The interaction of READ x CEFT x SEX was also significant ($F = 6.30$; $p < .01$). Both male and female, field dependent, good readers had the shortest mean orientation time.

Mode of presentation also varied significantly with the combination of both READ x CEFT ($F = 3.69$; $p < .05$) and READ x SEX ($F = 5.17$; $p < .05$). Shorter orientation times were discovered for field dependent, good readers in the dynamic condition, while female good readers also exhibited the shortest times in the dynamic condition.

Fixation

Significant differences in percentage of fixation on target between field independent and field dependent subjects were discovered in one segment ($F = 4.07$; $p < .05$) where field independents were characterized as having more fixations on target. Data for mode of presentation also confirmed the hypothesis on another segment for both ALL ($F = 4.57$; $p < .05$) and INIT ($F = 6.34$; $p < .05$) data. The percentage of fixations on target were not substantially differenct across modes for the other segments and averaged over the segments. Significant differences did occur between male and female on one segment (different than the above two segments) for ALL ($F = 4.38$; $p < .05$) and INIT data ($F = 4.80$; $p < .05$). In both cases girls fixated on target words significantly more than did boys.

One segment uncovered a READ x CEFT x PRES interaction (ALL - $F = 5.66$; $p < .05$) with the combination of poor reader x field independent x dynamic, demonstrating the largest percentage of fixations on target. Furthermore, two-way READ x PRES ($F = 4.83$; $p < .05$) led to significant findings when data was averaged for the segments taken together and only on INIT data.

Time

Data for two different segments (one for ALL data $F = 5.75$; $p < .05$ the other for INIT data $F = 5.57$; $p < .05$) resulted in significantly different percentages of time on target between good and poor readers. Both results supported the hypothesis. On examining Mode-Average data the hypothesis calling for reading proficiency differences on this measure was not confirmed.

Significant differences between the dynamic and static segments were revealed for ALL data on one segment ($F = 8.65$;

$p < .01$). In this particular case it was the static rather than dynamic presentation that resulted in more time spent on the target words.

Mode-Average data, on examining whole segments, resulted in significant male/female times ($F = 4.95$; $p < .05$). Girls spent a significantly greater percentage of time on targets than did boys.

CEFT x PRES was significant for one segment ($F = 10.08$; $p < .01$) with the combinations of field-independent static and field-dependent dynamic resulting in greater percentages of time on target.

Left-To-Right Directionality (L→R)

Significant results in an opposite direction than that predicted were found in early data for one particular segment ($F = 11.42$; $p < .01$). As well, early data for the same segment revealed that there was a significantly larger percentage of left-to-right movement in the dymanic presentation ($F = 8.65$; $p < .01$).

Average Duration of Fixation

In both ALL data for one particular segment ($F = 4.46$; $p < .05$) and averaged across segments ($F = 5.18$; $p < .05$) field independent subjects possessed larger fixation durations. Early data for two different segments showed that the static as opposed to the dynamic presentation did indeed lead to larger fixation durations ($F = 6.35$; $p < .05$) in one segment, while boys' fixations on target were significantly larger than girls' in another segment ($F = 6.60$; $p < .05$). ALL data on one particular cartoon segment lead to significant CEFT x PRES ($F = 5.54$; $p < .05$) and READ x PRES ($F = 3.83$; $p < .05$) interactions. CEFT x PRES was also significant for Mode-Average data (ALL - $F = 3.80$; $p < .05$). In both cases the shorter fixation durations were characteristic of field independent subjects viewing the static condition.

Discussion

The following is an interpretation of the results as they relate to the more general research objectives asked of this study.

1. By means of eye movement data, to gain more insight into the relationship between field articulation and reading ability.

Past research has indicated that field-independence-dependence was not closely related to the verbal component of reading. From the perceptual analysis performed here, there was very little evidence to suggest that field-independence-dependence was related to the perceptual component of reading proficiency (at least on the eye movement measures used to describe reading proficiency in this research). Although there may be communality on other measures between those leaning toward high reading proficiency and field independency, eye movement data in this study revealed little evidence to demonstrate that good readers were good readers because they possessed a field independent cognitive style. Naturally such an influence is limited to the eye movement data in this particular study, but it does raise the question as to what is the link between the two variables.

2. By means of eye movement research, to determine the effects of the media element of movement on the visual patterning of good and poor readers and field independent and field dependent individuals.

The media element studied here was the dynamic versus static quality of the medium of television. It had been purported in the literature with little supportive research that the dynamic qualities of television were a useful aid in attracting and maintaining attention to the screen. Gross (1974) and Allen (1975) believed in the essential information - relaying nature of the code of instruction. Studying the dynamic versus static quality was essentially an investigation into the code of the medium of television.

It appears from the data that in a rather complex stimulus presentation the dynamic moda did prove to be beneficial in the case of a few of the eye movement measures while in another more slow-paced animated segment the static presentation lead to more time being spent on the target words. In another segment in which sentences were formed in a left-to-right manner by means of animation, left-to-right scanning patterns were reinforced.

Generally, the visual cueing characteristic of dynamism within the medium of television did not result in many significant differences in viewing patterns as measured by the eye movement indices of this research. The auditory component of the stimuli may have diminished differences, for as Mock (1975) indicated, the auditory channel is a very strong cue as to where to look.

The general lack of significant findings in interaction with presentation mode seems to negate the possibility of visual supplantation. The results of this research suggests that movement was not necessarily an essential aid to particular reading or field-independent-dependent groups. Information as to where significant differences and interactions did in fact occur as a result of movement or lack of it should be exploited as a means of instructional assistance to particular viewers.

3. By means of eye movement research, to further investigate differences in perception between good and poor readers.

The results suggest that readers were only differentiated in the presence of particular qualities of a stimulus segment. As Rayner (1975) and Conant (1965) believed, it would appear that text not individual hypothesis testing determined eye movement patterns.

Questions which come to mind as a result of the analysis include the following: Was supplantation responsible for matching the eye movements of poor readers with those of good readers? Was it good readers' "sampling" (Wiener and Cromer, 1967) of the target that resulted in similar eye movements to those of poor readers in many instances? Were the stimuli not exciting enough to hold the attention of good readers? Were some of the segments not visually complex enough to allow for differences between the two groups ? Research questions such as these could be examined by strictly producing segments geared toward such an investigation.

The results of this eye movement study did not support previous claims that good readers exhibit shorter durations of fixation when reading. In more cases than not, they not only paused as long when on target but also spent more time on the words than did poor readers. The findings related to the effects of dynamic versus static stimuli also seem to weaken the supplantation argument as it relates to readers.

The results of this study point toward stimulus-specific qualities of particular segments that resulted in any differences between the two reader groups. Further analysis into isolated elements of these segments would yield information as to the possible occurrence of common elements that differentiated the groups. Knowledge such as this would be useful to instructional media designers.

4. Utilizing eye movement research to examine the perceptual differences between field independent and field dependent persons.

Similar to the results related to reading proficiency, there existed rather segment-specific eye movement differences involving field-independence-dependence. In one particular segment and for Mode-Average data, field dependents did pause significantly longer when on target as they made an attempt to decode the words. These results were supported in the literature by claims of cognitive delay and the global field dependent style of processing, while the shorter durations were more characteristic of the focusing in analytic mode of field independent individuals.

The question arises here as to whether the lack of significant differences is caused by individual or stimulus properties. Are qualities of the stimulus supplanting operations for field dependents or boring field independents? Are the stimuli not embedding enough to differentiate between the groups?

The shorter durations of fixation involving field independent subjects (at least for one segment and for Mode-Average data) seem to support a field independent sampling process that suggests that, characteristic to their style of processing information, field independent individuals spend time comparing target with non target areas as they attempt to find meaning in a stimulus complex.

It appears from this study that the task and its inherent qualities were what made light of the cognitive style differences individuals possessed. Further study of these stimulus-specific elements is one direction where future research could take aim.

Future Research

Of the questions posed by this research, the one which stands out is that which asks whether eye movements are stimulus or person-specific. The research performed here did lend insight into the problem as main effects and interactions were examined. The data indicates that individual qualities of each segment played a large part in where the subjects were looking. Further study of these stimulus specific elements is one direction where future research could take aim. As Olson (1974) and others suggested, we must research in depth the code or syntax of particular medium. More knowledge should be ascertained as to the effects of specific cueing devices, particularly auditory cues. Research comparing EMs, eliminating the voice-over, would no doubt yield worthwhile information as to the effects of the visual cues. The auditory component may have been the sole channel from where some of the subjects received information that assisted them in analyzing the field.

From the results of this study it appeared that in more cases than not the dynamic stimulus did not make a difference in the eye movements. If an asymptotic level was reached where the dynamic stimulus did not serve as a cueing aid, what was that level? Further research ought to concern itself with such a question. In other words, was there a level at which dynamism ceased to be beneficial and became an information processing hindrance?

Further study of EMPs as they relate to cognitive style variables ought to be performed. Resultant information would be useful to instructional designers and those researchers exploring the area of cognitive style. Other organismic variables including anxiety, stage of development (à la Piaget or Bruner), or locus of control could be included in such investigations.

Although many studies exist in this area, complete information concerning eye movements and reading has not been totally uncovered. Recent research (Mitrani, in Edelfat, 1975) has suggested that good readers "read" during saccadic EMs. Information such as this is useful in interpreting results although further exploration is needed in this area to substantiate his findings.

Not specifically compared, varying results were acquired between ALL and INIT data. What happened to the attention of subjects as they viewed a segment over time was a factor that must be taken into consideration when studying individual differences and different modes of instructional presentations. Further research into the most beneficial length of time for the exposure of a stimulus presentation ought to be investigated keeping attentional factors and pacing of a segment in mind.

Finally, sex differences as they relate to eye movements, the medium and organismic variables ought to be investigated further. This study produced some interesting main and interactive sex-related effects. However, it was one of the few EM studies to have done so. Many organismic variables have been correlated with sex, but it is rare to find an EM study which not only differentiated the "looking" behaviors of males and females, but also correlated these EMPs with other measures.

Educational Implications

Implications resulting from EM data surrounding organismic variables as they stand alone and in interaction with other factors related to the medium of television are useful to those involved in instructional design. Application of theory into practice need not be made solely to the medium of television, but can also be applied to other methods of communication (including the teacher) that are available in the teaching-learning environment. Knowing that qualities of a stimulus field may evoke certain EM responses in individuals, with a further investigation into the elements of these stimuli, instructional packages can not only be designed so as to take advantage of these attention-getting and maintaining devices, but can also be paced according to the perceptual style of information processors. Although it was found that the dynamic mode produced positive effects for some individuals while negative results for others, generally there were no significant differences due to mode. Information such as this is useful to those who have only looked at movement on the screen as a positive characteristic of the medium. Instructional materials designed for certain individuals might best be produced avoiding what may prove to be hindering devices.

As the National Institute for Education Report (Gibbon et al, 1974) suggested, EM studies are a useful "real-time" indicator of the perceptual processes of certain individuals. Eye movement studies allow one to study the style of individual processors and to investigate, as in this study, the effects of these styles under the influence of certain environmental factors. There are few dependent variables that can produce such useful information as it relates to individual differences. Individual differences are a major concern to educators as they investigate "person" qualities interacting with various environments and instructional methods. Knowledge of information processing skill gained from EM studies is useful to teachers and producers of instructional television who set up teaching-learning environments while keeping in mind an instructional pace suitable to various learners.

References

Allen, W. Categories of Instructional Media Research. Viewpoints: Bulletin of University of Indiana, 1970, 46, 1-13.

Allen, W. Intellectual Abilities and Instructional Media Design, Audio Visual Communication Review, 1975, 23, 139-170.

Conant, L. A Study of Eye Movement Behavior in the Reading Performance of Mature Readers on Reading Selections of Increasing Difficulty. Dissertation Abstracts International, 1965, 25, (11), 7075-7076.

Conklin, R., Muir, W., & Boersma, F. Field Dependency-Independency and Eye-Movement Patterns. Perceptual and Motor Skills, 1968, 26, 59-65.

Conway, J. Forms of Information Representation and Cognitive Operations. Viewpoints, Bulletin of University of Indiana, 1970, 46, 155-163.

Dwyer, F. Effect of Method in Presenting Visualized Instruction. AV Communication Review, 1973, 21, 437-451.

Edelfat, A. Concerning Differences Between Good and Poor Readers. Journal of Experimental Education, 1975, 43, 90-93.

Fleming, M. Eye Movement Indices of Cognitive Behavior. Audio Visual Communication Review, 1969, 17, 383-398.

Fleming, M. Media Research Projects: Considerations from a Perceptual Point of View. Viewpoints, Bulletin of University of Indiana, 1970, 46, 119-136.

Fowles, B. A Preliminary Proposal for Verbal Report as an Index of Effectiveness of Television Instruction. Unpublished Doctoral Dissertation, Yeshiva University, 1973.

Gagne, R. Learning Theory, Educational Media and Individualized Instruction. *Educational Broadcasting Review*, 1970, 4, 49-62.

Gibbon, S. et Al. *Report of the Conference of Visual Information Processing Research and Technology to the National Institute of Education*, 1974.

Gould, J. & Schaffer, A. Eye Movement Patterns during Visual Information Processing. *Psychonomic Science*, 1965, 3, 317-318.

Gould, J. & Schaffer, A. Eye Movement Patterns in Scanning Numeric Displays. *Perceptual and Motor Skills*, 1965a, *20*, 521-535.

Gould, J. Pattern Recognition and Eye-Movement Parameters. *Perception and Psychophysics*, 1967, 2, 399-407.

Gould, J. and Dill, A. Eye-Movement Parameters and Pattern Discrimination. *Perception and Psychophysics*, 1969, *6*, 311-320.

Gould, J. Eye-Movements During Visual Search and Memory Search. *Journal of Experimental Psychology*, 1973, 98, 184-195.

Gross, L. Modes of Communication and the Acquisition of Symbolic Competence. *N.S.S.E. Yearbook*, Chicago, 1974.

Guba, W., Wolf, W., deGroot, S., Knemeyer, M., Van Atta, R. & Light, L. Eye Movements and T.V. Viewing in Children. *Audio Visual Communication Review*, 1964, *12*, 386-401.

Jamison, D., Suppes, P. & Wells, S. The Effectiveness of Alternative Instructional Media: A Survey. *Review of Educational*

Research, 1974, 44, 1-67.

Mackworth, J. Some Models of the Reading Process: Learners and Skilled Readers. Reading Research Quarterly, 1972, 7, 701-733.

Mackworth, N. A Stand Camera for Line of Sight Recording. Perception and Psychophysics, 1967, 2, 119-127.

Mackworth, N.H. & Morandi, A. The Gaze Selects Information Details Within Pictures. Perception and Psychophysics, 1967, 2, 547-552.

Mackworth, N. & Bruner J. How Adults and Children Search and Recognize Patterns. Human Development, 1970, 13, 149-177.

Mock, K. The Relationship of Audio Visual Attention Factors and Reading Ability to Children's Television-Viewing Strategies. Unpublished Doctoral Dissertation, University of Toronto, 1975.

Nodine, C. & Lang, N. Development of Visual Scanning Strategies for Differentiating Words. Developmental Psychology, 1971, 5, 221-232.

Noton, D. & Stark, S. Scanpaths in Eye Movements in Pattern Perception. Science, 1971, 308-311.

O'Bryan, K. Eye Movements and Conservation. Unpublished Doctoral Dissertation, University of Alberta, 1969.

O'Bryan, K. & Boersma, F. Eye Movements, Perceptual Activity, and Conservation Development. Journal of Experimental Child Psychology, 1971, 12, 157-169.

Olson, D. Introduction. Media and Symbols: The Forms of Expression, Communication, and Education. N.S.S.E. Yearbook, Chicago, 1974.

Olson, D., Bruner, J. Learning through Experience and Learning through

Media. N.S.S.E. Yearbook, Chicago, 1974.

Rayner, K. The Perceptual Span and Peripheral Cues in Reading. Cognitive Psychology, 1975, 7, 65-81.

Rovet, J. Can Spatial Skills Be Acquired Via Film? An Analysis of the Cognitive Consequences of Visual Media. Unpublished Doctoral Dissertation. University of Toronto, 1974.

Salomon, G. & Snow, R. The Specification of Film Attributes for Psychological and Educational Research Purposes. Audio Visual Communication Review, 1968, 16, 225-244.

Salomon, G. Can We Affect Cognitive Skills Through Visual Media. Audio Visual Communication Review, 1972, 20, 401.

Salomon, G. What is Learned and How It is Taught: The Interaction Between Media, Message, Task, and Learner. N.S.S.E. Yearbook, 1974, 383-408.

Salomon, G. & Clark, R. Reexamining the Methodology of Research on Media and Technology in Education. Review of Educational Research, 1977, 47, 99-120.

Service, J. A User's Guide to the Statistical Analysis System. Student Supply Stores, Raleigh, North Carolina, 1972.

Snow, R. & Salomon, G. Aptitudes and Instructional Media. Audio Visual Communication Review, 1968, 16, 341-357.

Taylor, E. The Spans: Perception, Apprehension, and Recognition. As related to Reading and Speed Reading. American Journal of Opthamology, 4, 1957.

Taylor, S. Eye-Movement Photography with the Reading Eye. Educational

Developmental Laboratories, Inc. New York, 1960.

Tinker, M. Time Relations for Eye-Movement Measures in Reading. The Journal of Educational Psychology, 1947, 38, 1-10.

Vurpillot, E. Development of Scanning Strategies and Their Relation to Visual Differentiation. Journal of Experimental Child Psychology, 1968, 6, 632-650.

Wiener, M. & Cromer, W. Reading and Reading Difficulty: A Conceptual Analysis. Harvard Educational Review, 1967, 37, 620-643.

Witkin, H.A., Lewis, H.B., Hertzman, M., Machover, K., Meissner, P., & Wagner, S. Personality Through Perception. New York, Harper, 1954.

Witkin, H.A., Dyk, R.B., Faterson, H.F., Goodenough, D.R., & Harp, S.A. Psychological Differentiation, Studies of Development. John Wiley and Sons, Inc., New York, 1962.

Witkin, H.A. Origins of Cognitive Style. In Constance Scheerer (Ed.), Cognition, Theory, Research, Promise. New York, Harper and Row, 1964, 172-205.

Witkin, H.A. Oltman, P., Rasking, E., & Karp, S. Manual-Embedded Figures Test, Palo Alto: Consulting Psychologist Press, Inc., 1971.

Witkin, H., Moore, C., Goodenough, D., & Cox, P. Field Dependent and Field Independent Cognitive Styles and Their Educational Implications. Review of Educational Research, 1977, 47, 1-64.

Wolf, W. Perception of Visual Displays. Viewpoints, 1971, 47, 112-140.

IR007364

TITLE: PRINCIPALS' PERCEPTIONS OF ACTUAL AND IDEAL ROLES OF THE SCHOOL MEDIA SPECIALIST

AUTHORS: Ms. Sally Burnell
Media Specialist
Ames Community Schools
Ames, Iowa 50010

PRINCIPALS' PERCEPTIONS OF ACTUAL AND IDEAL ROLES OF THE SCHOOL MEDIA SPECIALIST

Introduction

School media centers have been undergoing changes in the last few years. One result of these changes is that school media centers have become the heart of the instructional program. The new functions performed by the school media specialist focus heavily on serving the needs of the professional staff and the needs of the learner. Because school media center activities are closely integrated with the educational program teachers, administrators and principals need to understand and support the center and the functions of the school media specialist. The most important non-media professional to the functioning of the media center is the principal. It is the principal's enthusiastic support, or lack of it, which often determines the success or failure of programs. If the principal supports the media specialists and their program the center has great potential. Without this interest and understanding, a significant media program is impossible.

Purpose of Study and Sample

Because the principal is a key force in the media center program, this study chose to describe the perceptions of the principal toward the role of the school media specialist. Two hundred high school principals in the state of Iowa were chosen to make up the sample. The randomized list of two hundred principals was obtained from the Iowa Department of Public Instruction.

This study focused on three general questions:

1) What perceptions do principals have toward the ideal roles and functions of the school media specialist?
2) What perceptions do principals have toward the actual performance of the school media specialist?
3) Is there a significant difference between the actual and ideal role of the school media specialist as perceived by secondary school principals?

In addition to studying the perceptions of principals as a whole, the principals were grouped according to the following variables and differences were studied:

1) Those who were employed in a small (enrollment of 500 or less) or large (enrollment of 501 or more) school.
2) Those who were employed in rural (population of 5,000 or less) or urban (population of 5,001 or more) geographical area.
3) Those who had a media course(s) or those who had not had a media course(s).
4) Those who had been employed in the school (tenure) for 0-5 years, 6-10 years, 11-15 years, or 16-more years.

The data collected was used to test the following hypotheses:

*1) There is no significant difference between the actual and ideal performance of the school media specialist as perceived by high school principals.

*2a) There is no significant difference between principals from small and large schools in their perceptions of what should be the actual role of the school media specialist.

2b) There is no significant difference between principals from small and large schools in their perceptions of what should be the ideal role of the school media specialist.

*2c) There is no significant difference between actual and ideal performance of the school media specialist as perceived by principals from small schools.

*2d) There is no significant difference between actual and ideal performance of the school media specialist as perceived by principals from large schools.

*3a) There is no significant difference between principals from rural and urban schools in their perceptions of what should be the actual role of the school media specialist.

3b) There is no significant difference between principals from rural and urban schools in their perceptions of what should be the ideal role of the school media specialist.

3c) There is no significant difference between actual and ideal performance of the school media specialist as perceived by principals from rural schools.

*3d) There is no significant difference between actual and ideal performance of the school media specialist as perceived by principals from urban schools.

4a) There is no significant difference between principals who had and had not taken a formal media course(s) in their perceptions of what should be the actual role of the school media specialist.

4b) There is no significant difference between principals who had and had not taken a formal media course(s) in their perceptions of what should be the ideal role of the school media specialist.

*4c) There is no significant difference between actual and ideal performance of the school media specialist as perceived by principals who had taken a media course.

*4d) There is no significant difference between actual and ideal performance of the school media specialist as perceived by principals who had not taken a media course.

*5a) There is no significant difference between principals with different tenure in their perceptions of what should be the actual role of the school media specialist.

5b) There is no significant difference between principals with different tenure in their perceptions of what should be the ideal role of the school media specialist.

*5c) There is no significant difference between actual and ideal performance of the school media specialist as perceived by principals with 0-5 years tenure.

*5d) There is no significant difference between actual and ideal performance of the school media specialist as perceived by principals with 6-10 years tenure.

*5e) There is no significant difference between actual and ideal performance of the school media specialist as perceived by principals with 11-15 years tenure.

Instrument Chosen for Data Collecting

The questionnaire was selected to be the type of instrument used for this study. The questionnaire was divided into two parts, each having its own purpose. The purpose of Part One was to obtain background

---

* Hypothesis was rejected at the 0.05 level of significance.

information on the principal and the principal's school. This background information included conditions that may have affected the principal's perceptions toward the actual and ideal role of the school media specialist.

The purpose of Part Two was to determine the principal's perceptions of school media specialist's performance (both the actual and the ideal performance). Part Two was composed of statements describing functions that could be performed by a school media specialist. The principal read each statement and indicated: a) his/her perception of the actual performance by the school media specialist of the function stated, and b) his/her perception of the level of performance by the school media specialist that would be ideal for that function. The principal indicated his/her perceptions by giving each statement a numerical value. A 0-5 scale was used to measure the principal's perception in regard to each statement. The content of part two was based on the recommendations from the Certification and Standards Committee of the Iowa Educational Media Association. Thus, Part Two of the questionnaire contained forty-one functions statements. These forty-one statements were later (during the statistical analysis) grouped into eight subroles or categories: systems, administration, selection, utilization, production, leadership, research, and general.

Testing of Hypotheses

When the respondents returned the questionnaire, the responses were categorized according to their corresponding subrole. For example, each questionnaire was taken and the responses for statements 1, 17, 18, and 31 were recorded. These statements defined the subrole systems. These four scores were then averaged in order to produce one score for the subrole "systems." Each questionnaire produced sixteen scores; there were eight subroles and two conditions (actual and ideal) under each subrole. The data was grouped according to this method so that perceptions of subroles instead of individual functions could be studied. The hypotheses were then tested by the paired t-test, t-test or analysis of variance. Significance was set at the 0.05 level.

Results

The test results allowed for the rejection of fourteen of the nineteen hypotheses. On the basis of the findings in this study, the following conclusions were made:

1) High school principals perceived a significant difference between the actual and ideal performance of the school media specialist. Significant differences were found in all eight subrole areas.
2) Each group of principals perceived a significant difference between the actual and ideal performance of the school media specialist. The two groups that were the most satisfied with the specialist's performance were: a) those principals employed in urban schools, and b) those principals employed eleven or more years in the school. The two groups that were least satisfied were: a) principals employed in small schools and b) principals employed in rural schools.
3) Principals from small and large schools perceived differently the actual performance of the school media specialist. Differences were found in all but one subrole-utilization. The two groups of principals perceived similarly the ideal performance of the school media specialist.

4) Principals from rural and urban schools received different scores in their perception of the actual role of the school media specialist. The significant differences between the two groups were found in the subroles: administration, selection, production, leadership, and general. In all significant cases, principals from urban schools produced the higher mean value. The two groups of principals did perceive similarly the ideal role of the school media specialist.

5) Principals who had taken a media course and principals who had not taken a media course did not differ on their perceptions of the actual role of their school media specialist. Neither did they differ on their perceptions of the ideal role of the school media specialist.

6) There was a significant difference between the perceptions of principals with different tenure toward the actual roles of the school media specialist. Principals employed 16 years or more produced a higher mean score than principals who have been employed 0 to 5 years. That is to say, principals employed 16 years or more perceived the media specialist performing the subroles more frequently than did principals employed 0-5 years.

Principal's Perceptions of Actual and Ideal Roles

A ranking of actual and ideal roles was produced from the principals sampled. On the actual level, the performance of the school media specialist in the various subrole categories ranged from performs infrequently (2.3) to performs sometimes (3.5) on a 5 point scale. Utilization and selection were the subroles being assumed most frequently; production and research were the subroles perceived as being least assumed by the school media specialist.

At the ideal level, the desired performance ranged from performs sometimes (3.2) to performs always (4.2) for the different subroles. Selection and utilization were the subroles that should be assumed most frequently by the school media specialist. Research and production were the subroles that should be performed least by the school media specialist.

The subroles were ranked almost identically in both the actual and the ideal charts. Principals were saying that the order in which the subroles were being performed by the school media specialist was correct. The desire of the principals was that more time needed to be spent on the higher priority subroles. The only subrole that differed by more than one in the rank ordering was the subrole general (which contained more traditional functions). Principals indicated that the specialists need to concentrate more effort in the area of administration and systems.

Actual/Ideal as Perceived by all Principals

There is a difference between the actual and ideal performance of the school media specialist as perceived by principals. None of the subroles were reported as being performed at the ideal level. It is very common in any field not to have programs functioning at the ideal level. Realities such as high cost, poor physical facilities, lack of professional help, etc. may place a barrier between the actual and ideal performance levels. However, the value of the results of this study was in its showing the subroles that had the greatest difference between the two conditions. In other words, the difference between actual and ideal performance. The widest gap existed in the administration subrole. More effort seems to be needed from the school media specialist in developing and implementing media programs. The two other subroles that needed the most improvement were systems and research. Generally the principals seem to desire that a school media specialist should: a) participate more as a member of the educational team; b) be more involved with determining goals for the media program; and c) apply the principals of research to the media program more frequently. Media specialists need to shift more of their efforts toward the learner.

It is this researcher's opinion that the gap between actual and ideal conditions under the subrole of research is due to the fact that research was not a part of, or not strongly emphasized in past media programs.

Conditions Effecting Actual and Ideal Perceptions

Do certain conditions effect the principal's perception toward the actual role and ideal role of the school media specialist? Considering perceived actual performance, differences between principals occurred for three of the four variables: a) population of the school, b) geographical location of the school, and c) tenure of the principal. The variable that produced the greatest difference between principals' actual perceptions was the size of the school. Seven out of eight subroles differed significantly (principals from large schools producing the higher score). It is assumed by this researcher that these results might have been due to different budgets, facilities, and the number equality of personnel working in large school media programs.

No significant differences for the ideal performance variable were produced for any of the four conditions or variables. Groups of principals did not differ in their desires of how the school media specialist should perform. This indicated that there was some unity within the sample of principals. Discovering that principals desired similar performance of the school media specialist can be valuable in program planning, developing in-service, educating future specialists, locating subroles where perhaps perceptions need to be changed, etc.

Recommendations

The data produced by this study indicated definite needs that should be examined in order to strengthen school media programs. Following are several recommendations towards fulfilling these needs. First, professional media staffs need to openly recognize the gaps between actual and ideal conditions in their schools. These gaps should be clearly identified so they may be acted upon. The staff members then need to discover what conditions were causing the gaps. After the identification process is completed, the staff then needs to plan short and long term activities that will close the gaps.

The widest gaps discovered in this study occurred in the areas of administration in systems, and in research. Since the principals surveyed indicated a willingness for the school media specialist to assume greater responsibility in these areas, the school media specialist should reassess their current activities and prepare to provide more input with regard to these three subroles. Second, school principals should seek to improve communication within the school. The principal should attempt to encourage other staff members to become aware of the functions of the school media specialist and to use the services of this individual.

Third, there needs to be an identification of the "new" functions of the school media specialist and the "traditional" functions. Principals ranked the traditional subrole (general) slightly above three other subroles.

Colleges and universities that train professional educators should make an effort to inform prospective school personnel about the duties and responsibilities of the school media specialist in order to change the image of the traditional library services and to diminish the tendency to perpetuate the stereotyped role of the school media specialist (historically known as the librarian).

Fourth, there was a substantial difference between the perceptions of principals employed in larger schools (higher student enrollment) and principals employed in smaller schools. Also, a significant difference occurred between principals in rural areas and urban areas. Again, staff members must identify the reasons for the discrepancies between actual and ideal conditions and strive to improve conditions for rural and small populated schools. Perhaps through discussion, in-service education, pooling together outside resources, etc. some of the actual conditions can come closer to the ideal. Curriculum planners for pre-service school media education should provide programs that recognize and address such differences as those found between rural and urban schools; large and small schools.

REFERENCES

American Association of School Librarians. Standards for school library programs. Chicago: American Library Assn., 1960.

American Association of School Librarians, Certification of School Media Specialists Committee. Certification model for professional school media personnel. Chicago: American Library Assn., 1976.

American Association of School Librarians, and Association for Educational Communications and Technology. Media programs: District and schools. Chicago: American Library Assn., 1975.

American Association of School Librarians and Department of Audiovisual Instruction, National Education Association. Standards for school media programs. Chicago: Library Assn.; Washington, D.C.: National Education Assn., 1969.

American Library Association, and National Education Association. Standards for school media programs. Washington, D.C.: American Library Assn., 1969. (ERIC Document Reproduction Service No. ED 031 086).

Anderson, C. J. Role expectations of the high school librarians as perceived by librarians, principals, and teachers (Doctoral dissertation, University of Oregon, 1970). Dissertation Abstracts International, 1971, 31, 5437A. (University Microfilms No. 71-10, 684).

Ball, H. G. A survey of issues, problems, and concerns affecting school media specialists in the southeast part of the United States. Normal, Alabama: Alabama A and M University, 1975. (ERIC Document Reproduction Service No. ED 103 028).

Bender, D. R. Services of the school media program. Baltimore, Maryland: Maryland State Dept. of Education, 1976. (ERIC Document Reproduction Service No. ED 126 908).

Biddle, B. J., and Thomas, E. J. Role theory: Concepts and research. New York: Wiley, 1966.

Bowden, M. G. The principal and the school librarian. Wilson Library Bulletin, 1963, 36, 572-577.

Brown, J. W., Norberg, K. D., and Srygley, S. K. Administering educational media (2nd ed). New York: McGraw-Hill, 1972.

Bucher, K. T. Role expectations held by professional school personnel for the roles of the school library media specialist (Doctoral dissertation, Auburn University, 1976). Dissertations Abstracts International, 1976, 37, 2467A. (University Microfilms No. 76-25, 689).

Cantor, P. F. Role expectations for library media services held by library media specialists, school administrators, and teachers (Doctoral dissertation, Columbia University, 1975). Dissertation Abstracts International, 1976, 36, 7707A. (University Microfilms No. 76-12, 728).

Chisholm, M., and Ely, D. P. Media personnel in education: A competency approach. Englewood Cliffs, New Jersey: Prentice-Hall, 1976.

Dale, E. Many things we see; and some of them we are. AV Instruction, 1964, 9, 64-67.

Davis, K. Human behavior at work: Human relations and organizational behavior (5th ed.). New York: McGraw-Hill, 1972.

Ellsworth, R., and Wagener, H. D. The school library: Facilities for independent study in the secondary school. New York: Educational Facilities Laboratory, 1963.

Faulkner, W. The town. New York: Random House, 1957.

Foskett, J. M. The normative world of the elementary school principal. Eugene, Oregon: The Center for the Advanced Study of Educational Administration, University of Oregon, 1967.

Frazier, A. Curriculum changes and the librarian. Wilson Library Bulletin, 1965, 39, 389.

Gaver, M. V. Services of secondary school media centers, evaluation and development. Chicago: American Library Assn., 1971.

Gerlach, V. S. The professional education of the media specialist. AV Communication Review, 1966, 14, 185-201.

Gibbs, W. F. The administration views the role of the librarian. Illinois Librarian, 1964, 45, 263-265.

Grimes, G. H. Audio-visual/library science relationships: A concept model and implication. Educational Technology, 1974, 14, 44-46.

Hall, R. O., and Harclevoad, F. F. The concent and pattern for the professional training of audiovisual communication specialists. Hayward, California: State College of Hayward, 1964. (ERIC Document Reproduction Service No. ED 003 780).

Hamreus, D. G. Media Guidelines. AV Instruction, 1970, 15, 31-34.

Harcleroad, F. F. (ed.). The education of the AV communication specialist. AV Communication Review, 1960, 8, 3-96.

Hughes, R. G. Jr. Media service cone. AV Instruction, 1970, 15, 31-34.

Martin, A. M., and Stone, C. W. A study of regional instructional media resources, Phase I -- Manpower. Pittsburgh, Pa.: University of Pittsburgh, 1965. (ERIC Document Reproduction Service No. ED 003 163).

McMahan, M. Building Coordinator, professional partner? AV Instruction, 1963, 8, 662-665.

Nelson, E. N. The media industry. Library Journal, 1970, 95, 1159-1161.

Olson, L. E. Teachers', principals', and librarians' perceptions of the school librarian's role. (Doctoral dissertation, University of Minnesota, 1966). Dissertation Abstracts, 1966, 27, 1846A (University Microfilm No. 66-12, 230).

Peterson, G. T. *The learning center: A sphere for non-traditional approaches to education*. Hamden, Conn.: Linnett Books, 1975.

Pfister, F. C., and Alexander, K. *Discrepancies between actual and ideal roles and functions of Texas school librarians as perceived by school superintendents, principals, and librarians*. Denton, Texas: Northern Texas State University, 1976. (ERIC Document Reproduction Service No. ED 134 188).

Popham, W. J., and Sirotnik, K. A. *Educational statistics: Use and interpretation* (2nd ed.). New York: Harper and Row, 1973.

Prostano, E. T. and Prostano, J. S. *The school library media center* (2nd ed.). Littletown, Colorado: Libraries Unlimited, 1977.

Schwilck, G. L. The library needs the principal. *The Bulletin of the National Association of Secondary School Principals*, 1966, *306*, 6-9.

State of Iowa, Department of Public Instruction. *Plan for Progress ... in the media center*. Iowa: Dept. of Public Instruction, 1970.

Stroud, J. G. *PSES, Purdue self-evaluation system for school media centers; junior, senior, high school catalog*. Lafayette, Indiana: Purdue University, 1976. (ERIC Document Reproduction Service No. ED 126 914).

Swarthout, C. *The school library as part of the instructional system*. Metuchen, N.J.: Scarecrow Press, 1967.

Walch, D. B., and Brumbaugh, D. W. Toward professionalization in the media field. *School Media Quarterly*, 1975, *4*, 27-36.

Wallington, J. C. *Jobs in instructional media*. Washington: Association for Educational Communication & Technology, 1971.

APPENDIX A: SAMPLE QUESTIONNAIRE

STUDY OF ACTUAL AND IDEAL ROLES
OF SCHOOL MEDIA SPECIALISTS

The purpose of this questionnaire is to attempt to identify the ideal roles and functions along with the actual roles and functions of the school media specialist. In order to clarify the term 'school media specialist', the following definitions are provided:

School librarian- professional responsible for print materials that are housed in the school media center (library).

Audio-visual specialist- professional responsible for non-print materials and equipment that are housed in the school media center (library).

School media specialist- professional responsible for all of the materials (print, non-print, equipment) that are housed in the school media center (library).

****
Print: books, magazines, etc.
Non-print: films, slides, audio tapes, etc.
****

For purposes of this survey, a school media specialist may be considered:
a) one person responsible for all of the materials in the school media center.
b) a two person team: librarian plus audio-visual specialist.
If your school employs a two person team as described above, you will need to think of both persons when rating the statements in Part 2.

PLEASE COMPLETE THE SURVEY BY:

1. Checking the correct answer for the questions asked in Part 1.
2. Indicating for each statement your perception of the actual condition in your school. (Part 2, left hand scale)
3. Indicating for each statement your judgment as to the ideal condition for that function. (Part 2, right hand scale)

Thank you!

PART 1

1. Check the grade levels you are responsible for as a secondary principal:

____ 7-12
____ 8-12
____ 9-12
____ Other

2. The total student population of your secondary school is:

____ Under-250
____ 251-500
____ 501-750
____ 751-1,000
____ 1,001-1,250
____ 1,251-over

3. The area in which your school is located can be considered:

____ Rural (population: under-5,000)
____ City (population: 5,001-50,000)
____ Urban (population: 50,001-over)

4. How would you describe your school media center?

____ Seperate: one person is responsible for the print media, the other person is responsible for non-print media and equipment.
____ Unified: one person is responsible for all media.

5. The main professional person(s) in charge of the school media center is: (NOTE: see definitions on first page)

____ School librarian(s)
____ Audio-visual specialist(s)
____ School media specialist(s)

6. Have you ever taken a formal media course at a college or university?

____ Yes
____ No

**** In answering questions 7-9, count each staff member only once. ****

7a. The number of school librarians your school employs is:

____0
____1
____2
____Other

7b. The total work hours per week for all of the librarians are:

____Under-20 hrs.
____21-39 hrs.
____40-45 hrs.
____46-59 hrs.
____60-79 hrs.
____80-over hrs.

8a. The number of audio-visual specialists your school employs is:

____0
____1
____2
____Other

8b. The total work hours per week for all of the audio-visual specialists are:

____Under-20 hrs.
____21-39 hrs.
____40-45 hrs.
____46-59 hrs.
____60-79 hrs.
____80-over hrs.

9a. The number of school media specialists your school employs is:

____0
____1
____2
____Other

9b. The total work hours per week for all of the school media specialists are:

____Under-20 hrs.
____21-39 hrs.
____40-45 hrs.
____46-59 hrs.
____60-79 hrs.
____80-over hrs.

10. The number of years that you have been employed as a secondary school principal at your school is:

____1-5 yrs.
____6-10 yrs.
____11-15 yrs.
____16-over yrs.

ACTUAL PERFORMANCE							IDEAL PERFORMANCE					
DON'T KNOW	NEVER PERFORMS	PERFORMS INFREQUENTLY	PERFORMS SOMETIMES	PERFORMS FREQUENTLY	PERFORMS ALWAYS		DON'T KNOW	NEVER PERFORMS	PERFORMS INFREQUENTLY	PERFORMS SOMETIMES	PERFORMS FREQUENTLY	PERFORMS ALWAYS
0	1	2	3	4	5	22. Participates in the planning, arrangement, and utilization and development of media facilities which will support the objectives of the media program and the instructional program.	0	1	2	3	4	5
0	1	2	3	4	5	23. Operates production equipment.	0	1	2	3	4	5
0	1	2	3	4	5	24. Evaluates and selects production equipment and supplies.	0	1	2	3	4	5
0	1	2	3	4	5	25. Assists teachers and students in identifying, obtaining, and adapting media to meet their needs.	0	1	2	3	4	5
0	1	2	3	4	5	26. Provides guidance in reading, listening, and viewing experiences for students and teachers.	0	1	2	3	4	5
0	1	2	3	4	5	27. Keeps school media center quiet.	0	1	2	3	4	5
0	1	2	3	4	5	28. Keeps an adequate equipment inventory.	0	1	2	3	4	5
0	1	2	3	4	5	29. Retrieves, interprets, and evaluates available research related to the educational program.	0	1	2	3	4	5
0	1	2	3	4	5	30. Provides and protects the right of access to information by faculty and students within the existing legal framework.	0	1	2	3	4	5
0	1	2	3	4	5	31. Determines goals for the media program as an integral part of the educational program of the school and district.	0	1	2	3	4	5
0	1	2	3	4	5	32. Establishes procedures for effective and efficient acquisition, processing, cataloging, circulation, distribution, and maintainance of materials.	0	1	2	3	4	5
0	1	2	3	4	5	33. Establishes job specifications and applies principles of personnel management.	0	1	2	3	4	5
0	1	2	3	4	5	34. Develops and applies criteria for evaluating and selecting a variety of materials and equipment.	0	1	2	3	4	5
0	1	2	3	4	5	35. Establishes and administers processes and procedures for preview, evaluation, selection, and withdrawal of materials and equipment.	0	1	2	3	4	5
0	1	2	3	4	5	36. Instructs and supervises others in media design and production.	0	1	2	3	4	5
0	1	2	3	4	5	37. Designs and adapts an identified research study for the development and advancement of the media program.	0	1	2	3	4	5
0	1	2	3	4	5	38. Delivers needed equipment to the classrooms.	0	1	2	3	4	5
0	1	2	3	4	5	39. Teaches skills in the retrieval and utilization of materials and equipment to students and teachers.	0	1	2	3	4	5
0	1	2	3	4	5	40. Designs production facilities and establishes basic routines for the operation of those facilities.	0	1	2	3	4	5
0	1	2	3	4	5	41. Designs, develops, and writes, proposals for the acquisition of local, state, and federal funds to support and expand media programs.	0	1	2	3	4	5

In order to identify those surveys that have been returned, PLEASE sign:

YOUR NAME: ______________________________

YOUR SCHOOL: ______________________________

THANK YOU!

PART 2

Please circle the number which best represents your perception of the actual and ideal performance of the media specialist for each of the statements:

ACTUAL PERFORMANCE							IDEAL PERFORMANCE					
DON'T KNOW	NEVER PERFORMS	PERFORMS INFREQUENTLY	PERFORMS SOMETIMES	PERFORMS FREQUENTLY	PERFORMS ALWAYS		DON'T KNOW	NEVER PERFORMS	PERFORMS INFREQUENTLY	PERFORMS SOMETIMES	PERFORMS FREQUENTLY	PERFORMS ALWAYS
0	1	2	3	4	5	1. Participates, as a member of the educational team, in the design and continual development of the curriculum.	0	1	2	3	4	5
0	1	2	3	4	5	2. Sets up and operates equipment.	0	1	2	3	4	5
0	1	2	3	4	5	3. Conducts in-service media training for teachers.	0	1	2	3	4	5
0	1	2	3	4	5	4. Initiates, develops, implements, and evaluates policies for the operation of the media program.	0	1	2	3	4	5
0	1	2	3	4	5	5. Prepares, justifies, and administers the media program budget.	0	1	2	3	4	5
0	1	2	3	4	5	6. Builds a collection of bibliographic aids and tools and other sources to provide current reviews and information about materials and equipment.	0	1	2	3	4	5
0	1	2	3	4	5	7. Develops selection policies which meet curricular, informational, and recreational needs and conforms to the appropriate legal requirements.	0	1	2	3	4	5
0	1	2	3	4	5	8. Produces media for specified learning objectives.	0	1	2	3	4	5
0	1	2	3	4	5	9. Teaches skills in the retrieval and utilization of materials and equipment to students and teachers.	0	1	2	3	4	5
0	1	2	3	4	5	10. Repairs malfunctioning equipment.	0	1	2	3	4	5
0	1	2	3	4	5	11. Reproduces printed material.	0	1	2	3	4	5
0	1	2	3	4	5	12. Applies the principles of research to the development and advancement of the media program.	0	1	2	3	4	5
0	1	2	3	4	5	13. Engages in research and publication activities.	0	1	2	3	4	5
0	1	2	3	4	5	14. Participates in district, county, regional, state, and national organizations.	0	1	2	3	4	5
0	1	2	3	4	5	15. Checks shelves frequently to keep materials in order.	0	1	2	3	4	5
0	1	2	3	4	5	16. Determines classification system to be used for organizing materials.	0	1	2	3	4	5
0	1	2	3	4	5	17. Analyzes learner characteristics, such as various abilities, interests, needs, and learning styles.	0	1	2	3	4	5
0	1	2	3	4	5	18. Recommends alternative learning environments.	0	1	2	3	4	5
0	1	2	3	4	5	19. Maintains an effective public relations program which communicates the vital contribution of the media program to learning.	0	1	2	3	4	5
0	1	2	3	4	5	20. Assesses the current status of the media program in terms of district, state, regional, and national guidelines and to establish short and long range plans.	0	1	2	3	4	5
0	1	2	3	4	5	21. Identifies and interprets legislation which affects the media program.	0	1	2	3	4	5
							0	1	2	3	4	5

Abstract

The purpose of this study was to investigate instructional methodology related to the effectiveness of different instructional feedback strategies that can be used in regular classroom situations. Experimental conditions were created to investigate whether, in addition to mastery learning strategies, (a) student self-evaluation of their products using specified criteria or (b) student self-evaluation and instructor feedback on the accuracy of their evaluations would affect either the student's ability to use stated criteria to judge their own performances, or would actually improve student performance. Fifty-six undergraduate students were assigned to either the control group, the self-evaluation group, or the self-evaluation/feedback group to measure the effects of these variables.

Results indicate that the association between student self-evaluations and instructor evaluations on a final class project was negatively affected by student self-evaluation only, and positively affected by the combination of self-evaluation and instructor feedback on the accuracy of the student self-evaluations. Feedback was shown to be most effective in situations in which students had too much confidence in the quality of their materials.

These findings indicate that the combination of student self-evaluation and instructor feedback on the accuracy of those evaluations does affect the ability of students to use prespecified criteria to evaluate their own work during the developmental process. In addition, it positively affects the performance of students who overestimate the quality of their work.

The Effects of Self-Evaluation as an Instructional Feedback Strategy

Lou M. Carey
Arizona State University

Larry Israelite
Arizona State University

Richard F. Schmid
Concordia University

In an effort to increase the effectiveness of instuction, many universities advocate the use of "objective based" instructional programs. In courses in which objectives are used, students are told what performances are expected of them during the course and are often given the criteria by which these performances will be judged. It is assumed they will use the criteria given to evaluate the quality of their own performances and to alter performances which are not as they should be.

When mastery learning is the issue, the ability to self-evaluate takes on increased importance. If learners are unable to evaluate their own work, if they are unable to discriminate between correct and incorrect performances during the learning process, the results of instruction will be less than optimal. The question to be considered is whether students are able to use objectives and criteria to evaluate their own performances, and if so, whether the ability to self-evaluate improves with practice.

Experimental situations were created to investigate the effects of two independent variables: (a) learners' self-evaluation of their products using specified criteria and (b) learners' self-evaluation with instructor feedback of the accuracy of their evaluations; on two dependent variables: (a) learners' ability to use stated criteria for judging their own products, and (b) learners' performances on course objectives.

Review of Literature

Literature on systematically designed instruction and the variety of instructional strategies that can be employed to improve learning is abundant. For the purposes of this study, literature that contained descriptions of (a) systematically designed instructional procedures related to objectives, standards, and practice and feedback and (b) self-evaluation as an instructional strategy was reviewed.

Instructional strategies. The writers first explored literature on systematic instructional design procedures and the effects of using systematically designed instruction. Particular areas studied were performance objectives, performance standards, and practice and feedback.

A number of learning theorists state that learning will improve as a result of pre-specified objectives (Gagné & Briggs, 1974; Gerlach & Ely, in press; Kibler, Barker & Miles, 1970; Kibler & Basset, 1977; Mager, 1961). More research is needed on the development of strategies that will increase the ability of students themselves to use objectives in learning situations. One question about objectives which remains unanswered is whether students can effectively use only statements of objectives for maximum instructional effectiveness and, if not, what other instructional techniques preceeding instruction can be employed along with objectives to facilitate learning.

The specification of performance standards is widely accepted as an integral part of an instructional objective. When performance evaluation is judgmental, when there is no clearcut right or wrong performance, the specification of performance criteria becomes appropriate. When a simple right or wrong response is not possible, Dick and Carey (1978) suggest that instructional objectives should include a checklist of the types of behaviors which will be expected when the performance of students is judged. This should give students a clear understanding of the nature of the required performance. However, the degree to which learners are able to use criteria in shaping their behaviors has not yet been established (Carey, 1976).

The use of practice and feedback are universally accepted as important aspects of instruction. Both are regularly included in systematically designed instructional materials. Research on practice and feedback were reviewed in this study relative to self-evaluation or student evaluation of their own work. Kulhavy, Yekovich and Dyer (1976) have shown that, in programmed instruction, certain conditions maximize the effects of instructor feedback on learner performance. They found that feedback in programmed instruction is most effective when a learner believes a test response is correct when, in fact, it is incorrect. They say that, in this case, learner response confidence is incorrectly high. When a learner believes a response is incorrect, and it is, Kulhavy et al. state that feedback is not effective because learners do not understand either the subject matter being tested, the test question, or both.

In programmed instruction, feedback indicates to the learner whether a particular response to a program frame is correct or incorrect, and the effects of feedback are measured by a subsequent test covering the same material. In some instructional situations, the feedback given on one performance is expected to help the learner with future performances which are similar. When complex, multi-dimensional performances are given, learners' confidence in this performance can be measured by having them evaluate each aspect of their own performance. Self-evaluation becomes an indicator of learner response confidence. If instructor feedback is then given on the quality of learner self-evaluations, in addition to the performance of tasks to be learned, a double feedback condition exists. The effects of feedback on student performance should be noticed under these conditions regardless of learner confidence, especially when the feedback given on performance and on self-evaluations is instructional and causes students to attend more to objectives and standards for acceptable performance.

Self-evaluation. One method for determining the degree to which learners are able to use objectives and criteria statements to shape their performances is to ask learners to use them to evaluate their own work. Those who are able to evaluate their work using pre-specified criteria probably possess a clear understanding of the objective and criteria concerned. On the other hand, learners who are unable to use performance criteria to judge the quality of their work after its completion are probably unable to use the same criteria during its production. Learners who are unable to use criteria during performances would not be expected to perform as well as those who have the ability to apply statements of criteria when developing instructional products.

Clark (1938) found that students were able to evaluate their own performance on college level algebra, quantitative analysis, and chemistry problems. The correlation between student scores and instructor scores was .80. Bennent (1958) supports the argument that students are able to evaluate their own performances.

Both Clark and Bennent report that students gave themselves the same grade as their instructors in a high percentage of cases. Estimating a single grade, or single number on a 5-point scale seems to be an easier task than determining how closely several parts of a complex performance match the related performance criteria. When one performs in the field, grades are not a consideration, job performance is, and evaluations are based on how closely the actual performance approximates the expected performance.

The question of whether self-evaluation actually promotes learning remains unanswered. McEowen (1957) found that learning was unaffected by students' evaluation of their materials. In this study, however, performance had no effect on course grades, so students had little at stake. Had the performance affected the grades, student interest in learning might have been greater and the results might well have been different.

In a study involving sculpture, graphics, painting, and drawing skills, Fried (1965) found that although sculpture improved as a result of self-evaluation, the other skills studied did not. Noting the inconclusiveness of his own finding, Fried concluded that the value of self-evaluation was still undetermined, and recommended continued research in the area.

In a study conducted in two Air Force technical schools, Duel (1958) found that achievement was improved when students were given formal and periodic opportunities to evaluate their own work. A study in which self-evaluation had a positive effect on students' mechanical drawing skills was conducted by Irwin (1973). He found that students who evaluated their own materials throughout the school year learned more as measured on a standardized test which assessed mechanical drawing skills. In this study, however, results were confounded because the self-evaluating group had a significantly higher mean IQ than the control group.

Research into the effects of self-evaluation has generally been concerned with either student ability to self-evaluate, or the effect of self-evaluation on student performance. An indication from the literature is that, in certain cases, students are able to evaluate their own materials, although the evidence is far from conclusive. Evidence has been cited which both supports and contradicts the assertion that self-evaluation improves student performance. Little has been written on the effects of (a) student practice on evaluating their own work and of (b) instructor feedback on student self-evaluation. The combination of systematic instructional procedures with self-evaluation as an additional feedback strategy may prove to make instruction more effective.

Hypotheses

The literature concerning learner self-evaluation, though limited, does indicate that students may possess at least a limited ability to evaluate their own work and that periodic self-evaluation may improve learner performance.

The specific hypotheses tested in this research are the following:

1. The correlation between learner and instructor evaluation scores on a final course project will be higher when instructor feedback is given on previous learner self-evaluations throughout the term than when no instructor feedback on self-evaluation is given, or when students do not participate in self-evaluation throughout the term.

2. Learner performance on a final class project will be better when students evaluate their own materials throughout the term and receive instructor feedback on their self-evaluations than when students do not receive feedback on their self-evaluations, or do not participate in self-evaluation throughout the term.

3. Learner performance on a final product will be better when feedback is given on the quality of the students' evaluation of their own work than when instructor feedback is given only on the quality of the product for a subgroup of students who incorrectly assume that their initial products are well produced.

4. Learner performance on a final product will be better when feedback is given on the quality of the initial product and on student evaluations of the project, than when feedback is given only on the product for a subgroup of students who correctly assume their product is not well produced.

Pilot Study

In an attempt to identify those experimental procedures which could most easily be adapted to an ongoing instructional program, a pilot study was conducted during the spring semester of 1978 at Arizona State University, Tempe, Arizona.

Method

Sample. Three instructors, and nine undergraduate media production classes participated in the study. Three intact classes were assigned to each of the three treatment groups; control, self-evaluation, and feedback. Each instructor was responsible for only one treatment group.

Procedures. The control group used performance objectives, criteria statements, and instructor feedback to produce a mediated unit of instruction. There were three major checkpoints during the semester at which students formally submitted their materials for instructor evaluation and feedback. The self-evaluation group followed procedures previously described for the control group. In addition, prior to submitting their materials for instructor evaluation, they used the instructor evaluation form to assess the quality of their own work. Instructor feedback on the quality of the materials was the same for both the control and the self-evaluation groups. The third or feedback group produced their materials in the same manner as the other two groups. Like the self-evaluation group, prior to submitting their materials, they evaluated their own work and submitted the evaluation forms to their instructors. Instructor feedback for this group consisted of both feedback on the quality of the products and on the quality of the student evaluations.

To enable comparisons in self-evaluations among the three groups, the control group used the evaluation forms to assess their final products at the end of the semester.

Results. Pearson product moment correlations were used to compare the evaluation scores of students with those of their instructors in the control, self-evaluation, and feedback groups on the final product. Significant correlations ($p \geq .05$) were observed between student and instructor ratings in both the control and the feedback groups. A slightly negative, but non-significant correlation was observed in the self-evaluation group.

A one-way analysis of variance among group 5 revealed no significant differences ($p \geq .05$) among the performance scores of the three groups.

Discussion

As a result of the pilot study, a number of procedural problems were identified. Due to the nature of the procedural problems, data collected during this study is not valid, and it is not generalizable to other instructional settings. Changes in procedures which were identified as being problematic are described in the following paragraphs.

The first procedure which was found to be unacceptable was the assignment of one treatment to each instructor. Training sessions were held to establish scoring guidelines, and a high degree of inter-rater reliability was established. However, during the study, which spanned an entire semester, a distinct pattern of instructor bias developed. One instructor seemed to be more lenient in grading than did the others, and also offered students more assistance in the development of their products. Students were assured of high grades by the instructor prior to formal instructor evaluation. As a result, students knew their scores before submitting their products, and the relationship between student and instructor scores was no doubt inflated in this case. It was determined that if each instructor delivered all three treatments, then the effects of this type of instructor bias would be minimized.

A second problem which was encountered concerned the instruments used for both student and instructor evaluations. The objectives list and grading criteria were on separate forms and although students had both forms in their possession at the time of self-evaluation, they rarely referred to the criteria sheet when grading their products. The objectives and criteria were combined into one form to help students focus on the criteria when assessing the quality of their materials.

The pilot study helped refine experimental procedures and enabled the researchers to conduct the main study in an ongoing instructional setting. With the exception of instructor treatment assignment and evaluation instrument design, the procedures used in the pilot study were also used in the main study.

Method

Sample

The study was conducted during the 1978 summer session at Arizona State University, Tempe, Arizona. Student participants were enrolled in IME 411, Audiovisual Materials and Procedures in Education. A total of 56 preservice and inservice teachers in three class sections participated. Students in each section were blocked according to three levels of grade point average; low, average, and high, and then randomly assigned to either the control, first, or second treatment group. At the onset of the study, there were 20 students in the control group, 16 in the first treatment group, and 20 in the second treatment group. During the conduct of the study, eight students were dropped from the control group because their instructor confused their treatment. Each course instructor taught the control and the two treatment groups in order to avoid instructor bias.

Mastery instructional strategies including use of performance objectives, specified criteria, examples of acceptable products, relevant practice, and instructor feedback on the quality of students' products were the instructional techniques used in all three treatment groups to help students produce one mediated and one non-mediated unit of instruction. The instruction, practice and feedback, and assessment procedures occurred throughout the course. There were three major checkpoints at which students formally submitted their products for instructor evaluation and feedback.

The control group followed the instructional procedures outlined previously to produce their materials during the course. Students in the first treatment group followed these same procedures, and in addition, prior to submitting their materials for instructor evaluation at each checkpoint, students in this group used the instructor evaluation form to evaluate their materials. The self-evaluations were submitted with the materials and were not seen again by students. The evaluation form included a listing of all aspects of the products to be evaluated and the specific criteria to be used in the evaluation. (Sample instructor evaluation forms can be seen in Appendix A.) This treatment group is referred to as the self-evaluation group.

The second treatment group, referred to as the feedback group, followed the same procedures as the self-evaluation group with the addition of one procedure. When instructor feedback was given on the quality of student materials, students were also given feedback on their self-evaluations. Instructors asked students why they graded themselves the way they did, based on the criteria, and showed students how their evaluations compared with instructor evaluations.

Students in both the control and self-evaluation groups received one type of feedback -- feedback on the quality of their materials. Students in the feedback group received feedback on the quality of their materials and also on their self-evaluations. It is important to note that the instructors did not see the student self-evaluations until after student products were graded. This eliminated instructor bias caused by student perceptions of the quality of their work.

Performance Standards

Each aspect of student products was graded by course instructors on a scale of 1 to 10, with 10 being the highest score possible. Grading criteria were established by course instructors prior to the beginning of the study. Instructor scores were used as tlhe standard against which all comparisons were made.

Design

In this study a posttest-only control group design was used to measure the effects of student self-evaluation of instructional products, and instructor feedback on the quality of self-evaluation on student performance on the final class product. An analysis of variance at a significance level of .05 was used to measure the effects of these two independent variables. A Pearson product moment correlation at a significance level of .05 was used to measure the relationship between student evaluations and instructor evaluations of all instructional products.

Results

Self-evaluation and feedback on the accuracy of self-evaluation were studied to ascertain their effects on (a) student's evaluation of their own performance and (b) on student achievement.

Effects of Self-Evaluation and Feedback on the Relationship Between Student and Instructor Scores.

It was hypothesized in this study that the correlation between student evaluation scores and instructor evaluation scores on the final course project would be higher when instructor feedback on student evaluations of other class projects was given than when no instructor feedback was given, or when no student evaluation took place.

Student and instructor evaluation means and standard deviations for the final course project and the correlation between them are shown in Table 1. For the control group, the student evaluation mean was 57.9 (maximum score = 60) and the instructor mean was 56.82, yielding a difference of 1.08 points with the correlation between them being .68 ($p \geq .05$). For the self-evaluation group, the student mean was 58.5, with the instructors giving a mean score of 53.67, differing by 4.83 points. The correlation between these two scores was -.07, which was not significant. Students in the feedback group predicted a mean score of 55.70 while the instructors' mean equaled 54.75, yielding a difference of .95 points. The correlation between these scores was .81 ($p \geq .05$).

Insert Table 1 about here

Figure 1 contains an illustration of the correlations between student and instructor scores at the first and third check points for the self-evaluation and feedback groups. There was very little change in student ability to evaluate their own work in the self-evaluation group. At the first checkpoint, the correlation between student and instructor scores was -.048 and at the third checkpoint, -.076. Neither of these correlations was

Table 1

Student and Instructor Mean Performance Scores and Standard Deviations on Final Class Project and Correlation between the Two Means for Each Treatment Group

Group	Student n[a]	Instructor	Standard Deviation	Correlation
Control	57.90 (11)	56.82	3.65	.68*
Self-evaluation	58.50 (14)	53.67	5.57	-.07
Feedback	55.70 (20)	54.75	8.26	.81**

Note. Maximum score = 60.

[a] Numbers in parenthesis indicate the number of students in each group.

*$p < .01$.

**$p < .001$.

significant. There was little association at the outset of the study and only slightly more at the conclusion.

On the other hand, the correlation between student and instructor scores for the feedback group was .49 ($p < .05$) at the first evaluation checkpoint and .81 ($p < .05$) at the third. The relationship between the two sets of scores for this group increased approximately 65 per cent.

Insert Figure 1 about here

Effects on Performance

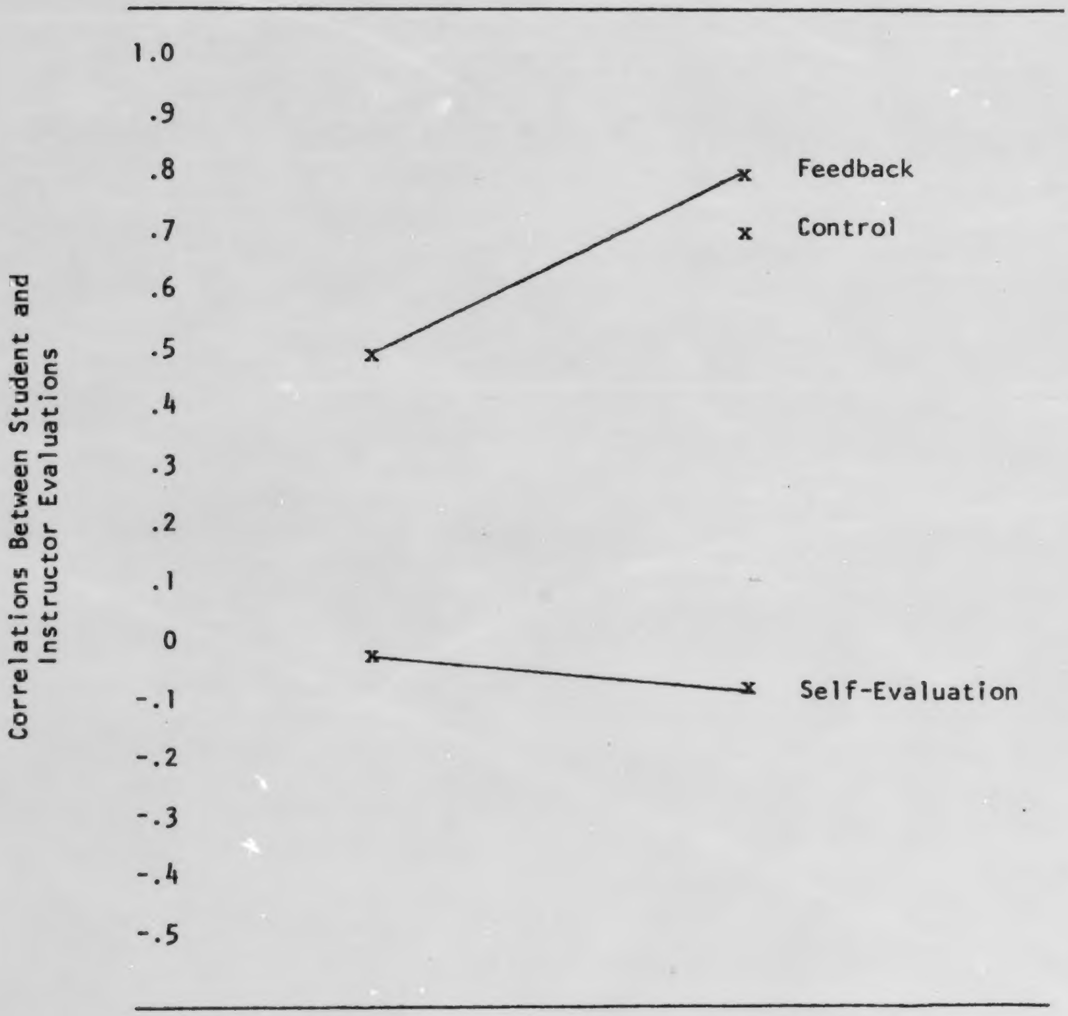
It was predicted that students' performance on the final class project would be better when students evaluated previous class products and received instructor feedback on the self-evaluations than when they completed self-evaluations alone or did not complete self-evaluations at all. A one way analysis of variance revealed no significant difference between group means.

To measure the relationship between student confidence in their materials as measured by their self-evaluation scores at the first evaluation checkpoint, and the effect of instructor feedback on student self-evaluations, students were blocked into two groups: (a) student evaluations above 50 indicating a high degree of confidence in their materials and instructor evaluations below 50 on the same materials, and (b) both student and instructor evaluations below 50 indicating low student confidence in their materials. Table 2 shows the means and standard deviations for these two subgroups for both the self-evaluation and feedback groups.

Insert Table 2 about here

Figure 1

Student and Instructor Evaluation Correlations
on the First and Final Evaluations



	First Evaluation	Final Evaluation
Control	--	.68
Self-evaluation	-.03	-.07
Feedback	.49	.81

Table 2

Instructor Mean Performance Scores and Standard Deviations on Final Project Blocked on Student Confidence in Their First Project

Group	Mean Score	n[a]	Standard Deviation
Incorrect high confidence			
Self-evaluation	53.09	(10)	5.65
Feedback	58.25	(12)	1.82
Correct low confidence			
Self-evaluation	55.25	(4)	5.85
Feedback	49.50	(8)	11.29

Note. Maximum score = 60.

[a] Numbers in parenthesis indicate the number of students in each group.

A one-way analysis of variance indicated that when student confidence was incorrectly high, there was a significant difference between the self-evaluation and feedback groups $\underline{F} = 9.03$ (1, 21) $\underline{p} \leq .01$. When student confidence in materials was correctly low, the feedback group mean was approximately six points lower than the mean score of the self-evaluation group.

Discussion

The purpose of this study was to investigate the effects of (a) student self-evaluation and (b) instructor feedback on the self-evaluations on the ability of students to evaluate their own materials and on their performance. Although criteria statements are often included in performance objectives, there has been little evidence which indicates that students can always use criteria to shape their performance (Carey, 76). The instruction used in this study included objectives with criteria statements and relevant practice and feedback. It conformed to the prescription for well developed instruction as specified by Gagné and Briggs (1974), Dick and Carey (1978), and Ely and Gerlach (in press).

The relationship between student score predictions and instructor scores on the final class product was reasonably high for the control group, $r = .68$ ($p \leq = .05$), supporting the work of Clark (1938) and Bennent (1969). However, the combination of self-evaluation and instructor feedback on self-evaluation raised this figure to .81, indicating that the skill of self-evaluation can be developed. These findings support the hypotheses stated in this paper. Criteria specific feedback on self-evaluation seems to be an effective way to develop student skills in estimating the quality of their performances.

The correlation between student and instructor evaluation scores for the control group was much higher at the third evaluation checkpoint than the correlation for either the self-evaluation or feedback group at the first checkpoint. For the control group the final product was their first opportunity to evaluate their own work. One possible explanation for this phenomenon is that as students go through a course, they acquire an understanding of instructors' expectations and how well their performances meet these expectations. Students then adjust their performances accordingly.

It was anticipated that self-evaluation itself without any instructor feedback on the accuracy of the evaluation would raise the correlation between student and instructor scores. The analysis revealed that self-evaluation itself had no beneficial effects, and in fact, seemed to have a negative effect. Self-evaluation seemed to confuse students and hamper their ability to use prespecified criteria to assess the quality of their materials. We have no explanation for the occurrence of this effect in either the pilot study or this study. There is no literature on which to base an explanation of this kind of result.

The effects of self-evaluation, and self-evaluation and feedback on student performance is difficult to gauge in this study. Students in all three groups achieved a grade of A on the final project scoring an average of 55 out of a possible 60 points. There was little variability in student performance.

When tied to student confidence in their materials, there was a significant difference between group means. Feedback on self-evaluation of earlier material seems to be highly effective in promoting improved student scores on later performances when student confidence in materials was incorrectly high. When students received instructor feedback on the quality of their materials and on their own evaluations, they had the opportunity to learn not only what instructors thought of their materials, but how their self-evaluations compared to instructor evaluations on an objective by objective basis. This technique pushed the mean scores of students in the unwarranted high confidence subgroup 3.5 points above the feedback group mean from 54.75 to 58.25. This tends to support the conclusions of Kulhavy, Yekovich & Dyer (1976) that feedback is most effective when a high degree of student confidence in materials is unwarranted. In the self-evaluation group, the difference between the mean scores of students who had incorrectly assumed their materials were good and the entire self-evaluation group was minimal, lending more support to the conclusion that the addition of feedback on the self-evaluations is beneficial when learner confidence in their work is incorrectly high.

When student confidence in materials was correctly low, feedback on self-evaluation seemed to have a negative effect on performance, lowering the subgroup means from 54.75 to 49.50. These results contradict the hypothesis stated earlier that feedback on self-evaluations is effective for all students, but support the findings of Kulhavy, et al. who concluded that when student confidence was correctly low, they were confused, and feedback was not effective. The results of this study indicate that when students were unsure of their work, additional feedback only caused more confusion.

Duel (1968) and Irwin (1973) both concluded that self-evaluation did improve student performance. Due to the ceiling effect noticed in this study, the current results can neither support their conclusions, nor refute the findings of McEowen, who found no improvement, and Fried (1965), who found improvement in some situations. It is clear that the effects of self-evaluation on student performance would possibly be more informative under conditions in which student grades are normally distributed.

Although the research seems to support the results of the Kulhavy, et al., (1976) research concerning the effects of feedback, continued research in this area is also indicated in order to clarify the effects of feedback in learning situations in which student performances are complex and student confidence is difficult to gauge.

There were many questions raised during this study that will need to be investigated in future studies. Two were of particular interest. The first question relates to the unexplained negative effects of self-evaluation only on student evaluations of their own work. The researchers expected this activity to increase the ability of students to assess their own work. When instructors appeared to ignore student evaluations, students were not helped.

A second question of interest is where to begin instruction for students who are unable to use prespecified criteria to improve the quality of their work. If students cannot perform specified tasks, cannot use objectives and criteria statements to evaluate their own work, and cannot use instructor feedback to improve their products, then the problem may be that these students do not possess prerequisite skills for the instruction. A feedback strategy that includes information on current instruction and on prerequisite skills instruction should be investigated in an attempt to identify the best strategy for this particular subgroup of students.

If students can learn to use prespecified objectives and criteria effectively to evaluate and shape their performances, the case for the use of instructional objectives is made even stronger. Then, not only will students know precisely what is expected, but they will be able to tell if and when they are able to perform the specified tasks.

If student self-evaluation can be developed to an acceptable level, then time spent on instructor feedback can be greatly reduced, freeing instructors for other instructional tasks during the educational process.

References

Bennett, K. B. Student evaluation -- the new way. *Industrial Arts and Vocational Education*, 1958, *47*, 300-302.

Carey, L. M. The investigation of self-evaluation procedures for identifying the instructional needs of teachers. (Doctoral Dissertation, Florida State University, 1976). Dissertation Abstracts International, 1977, *37*, 6428A. (University Microfilms No. 77-8575)

Clark, P. E. Can college students evaluate themselves? *School and Society*, 1938, *47*, 614-16.

Dick, W., & Carey, L. M. *The systematic design of instruction*. Glenview, IL: Scott, Foresman and Company, 1978.

Duel, H. S. Effect of periodic self-evaluation on student achievement. *Journal of Educational Psychology*, 1958, *49*, 1979.

Fried, H. C. Effect of end product self-evaluation on college students' art performance. (Doctoral dissertation, Illinois State University, 1965). *Dissertation Abstracts,* 1967, *27*, 17034. (University Microfilms No. 66-5370)

Gagné, R. M. & Briggs, L. J. *Principles of instructional design*. New York: Holt, Rhinehart and Winston, 1974.

Gerlach, V. S. & Ely, D. P. *Teaching and media -- a systematic approach* (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall Inc., in press.

Irwin, J. L. An investigation of the effects of student self-evaluation and marking on achievement of cognitive and affective objectives of a basic mechanical drawing course. (Doctoral dissertation, Pennsylvania State University, 1973). *Dissertation Abstracts International*, 1973, *34*, 1484A. (University Microfilms No. 73-24, 013)

Kibler, R. J., Barker, L. L., & Miles, D. T. Behavioral objectives and instruction. Boston: Allyn and Bacon, Inc., 1970.

Kibler, R. J. & Bassett, R. E. Writing performance objectives. In L. J. Briggs (Ed.), Instructional design, principles and applications. Englewood Cliffs, NJ: Educational Technology Publications, Inc., 1977.

Kulhavy, R. W., Yekovich, F. R. & Dyer, J. W. Feedback and response confidence. Journal of Educational Psychology, 1976, 68, 522-528.

Mager, R. F. Preparing instructional objectives. Palo Alto, CA: Fearon Publishers, 1961.

McEowen, R. H. A comparison of four different techniques for improving performance through student self-evaluation (Doctoral dissertation, East Texas State University, 1967). Dissertation Abstracts, 1967, 27, 4049A. (University Microfilms No. 67-7356)

APPENDIX A

IME 411
SS-1-78
Planning Document Evaluation

NUM ________
1 2 3

NAME ________

Directions -- Using a 1 to 10 scale, with 1 being the lowest and 10 the highest, evaluate each section of the planning document. Please use only whole numbers (no fractions). Base your evaluations on the criteria supplied for each section. Thank you. Each area has a blank to account for any other considerations which may come up.

(1-10) _____ 1 - Content Outline
- a) Is the content of your outline fully explained?
- b) Is the skill or concept being taught covered in its entirety?
- c)

(1-10) _____ 2 - Need - Purpose
- a) Is the statement of the purpose a clear statement of the teacher's goal?
- b) Is the need for the unit clearly described?
- c)

(1-10) _____ 3 - Audience Analysis
- a) Does the audience analysis contain some or all of the following:
  - 1) class size?
  - 2) description of ability level of students?
  - 3) description of racial makeup of the class?
  - 4) description of different socio-economic groups found in class?
  - 5)

(1-10) _____ 4 - Objectives
- a) Are objectives meaningful?
- b) Are objectives related to unit of instruction?
- c) Are objectives:
  - 1) student oriented?
  - 2) behaviorally stated?
- d) Do objectives:
  - 1) specify conditions of performance?
  - 2) specify standards of performance?
- e)

(1-10) _____ 5 - Media Selection
- a) Does media have potential for helping learners?
- b) Are there other more appropriate media which could make the instruction more effective or efficient?
- c) Are there any special circumstances that affect media selection which have not been considered?
- d)

(1-10) _____ 6 - Utilization
- a) Is the process for implementing the unit clear and sensible?
- b) Are all the materials utilized?
- c) Is the teaching strategy consistant with the audience, objectives, purpose, and time constraints?
- d)

IME 411
SS-1-78
Materials Evaluation

NUM ______________
1 2 3

NAME ______________________

Directions -- Using a 1 to 10 scale, with 1 being the lowest and 10 the highest, evaluate each aspect of the materials produced for the media product. Please use only whole numbers (no fractions). Base your evaluation on the criteria supplied for each aspect. Blanks are left to account for any other considerations which may come up. Thank you.

TECHNICAL QUALITIES CRITERIA

(1-10) _____ 1. Lettering
a) Is lettering legible?
b) Are letters neatly done?
c) Are letters and words appropriately spaced?
d) Are there any unwanted spots of ink or ink smears?
e) Have guidelines been completely erased?
f)

(1-10) _____ 2. Illustration
a) Is inking neat and accurate?
b) Is there sufficient contrast between original/board?
c) Have all pencil/stray marks been cleaned off?
d) Does enlargement fit board properly?
e)

(1-10) _____ 3. Mounting
a) Is visual completely adhered to backing?
b) Is visual trimmed neat and clean?
c) Have registration marks been erased?
d) Has excess MT-5, Sealamin, and rubber cement been removed?
e) Is visual well-positioned?
f)

(1-10) _____ 4. Transparency making
a) Are lines solid when projected?
b) Are colored areas solid and clean?
c) Is hinging secure and workable?
d) Is transparency neat and clean?
e)

(1-10) _____ 5. Construction
a) Is product constructed soundly or held together securely?
b) Has adequate caution been taken to assure product will stand up to classroom use?
c)

(1-10) _____ 6. Appearance
a) Is the product neat?
b) Have any errors been made which detract from the product's overall appearance?
c)

(1-10) _____ 7. Other (please list)
a)
b)

IME 411
SS-1-78
Mass Media Project Evaluation

NUM ____________________
1 2 3

NAME ____________________

Directions -- Using a 1 to 10 scale, with 1 being the lowest and 10 the highest, evaluate each aspect of the mass media product. Please use only whole numbers (no fractions). Base your evaluation on the criteria supplied for each aspect. Blanks are left to account for any other considerations which may come up. Thank you.

MASS MEDIA PROJECT CRITERIA

(1-10) _____ 1. Need-Purpose
- a) Is the statement of the purpose a clear statement of the teacher's goal?
- b) Is the need for the unit clearly described?
- c)

(1-10) _____ 2. Audience Analysis
- a) Does the audience analysis contain some or all of the following
  - 1) class size?
  - 2) description of ability level of students?
  - 3) description of racial makeup of the class?
  - 4) description of different socio-economic groups found in class?
  - 5)

(1-10) _____ 3. Description of Program
- a) Title
- b) Time
- c) Where can it be viewed?
- d) Length
- e) Medium (TV, Film, etc.)
- f)

(1-10) _____ 4. Synopsis of Program
- a) Is program clearly described?
- b) Is focus of lesson identified?
- c)

(1-10) _____ 5. Instructional Activities
- a) Do preprogram activities set up a framework for viewing?
- b) Is the relationship between preprogram activities, postprogram activities and the program itself clear?
- c) Do activities help learner attain the objective of the lesson?
- d)

(1-10) _____ 6. Objectives
- a) Are objectives meaningful?
- b) Are objectives related to unit of instruction?
- c) Are objectives:
  - 1) student oriented?
  - 2) behaviorally stated?
- d) Do objectives:
  - 1) specify conditions of performance?
  - 2) specify standards of performance?

An Empirical Analysis of the Characteristics of Performance Objectives*

Vernon S. Gerlach
Arizona State University

Richard F. Schmid
Concordia University

Purpose

The concept of instructional systems design is inextricably intertwined with the concept of performance objectives. The latter are referred to by many different terms, such as behavioral objectives, terminal learning objectives, criterion referenced objectives, and many others. In nearly every model for the design of instructional systems, the performance objective is the point of departure; decisions regarding measurement, evaluation, production, and sequencing (to mention only the obvious) are "objective-based."

Most advocates of performance objectives agree that an adequate objective consists of three basic components. Consider this performance objective: To measure the diameter of a solid cylinder, given a tubular micrometer, to the nearest .001." Note the three components: (1) the verb, which states what the trainee will be able to do after his training (to measure); (2) the condition of performance, which states the circumstances under which the action occurs (given solid cylinders to be measured and a tubular micrometer); and (3) the criterion, which states the standard of performance (to the nearest .001").

At three recent AECT conventions, we have reported the results of our research in the area of the behavioral characteristics of behavioral objectives. Basically, we have been attempting to find answers to the question, to what degree do the several components influence readers' perceptions of the observability of various performance objectives or of the components of such objectives? The findings reported have met generally accepted criteria for reliability. However, our data provided no basis for answering the question, to what degree do these findings apply to the area of military instructional systems design, as opposed to teaching and instructional development in public schools? A simple correlational study was designed to provide answers to this question.

Rationale

Deno and Jenkins (1969) found that many verbs which elementary and secondary school curriculum experts and instructional designers recommend for instructional planning were not perceived by educators to be "observable." They concluded that verbs used in behavioral objectives are selected for consistency of usage rather than for maximum observability. Gerlach (1974)

---

*This paper is based in part on research funded by U.S. Air Force Office of Scientific Research Grant #76-2900, awarded to Arizona State University. Vernon S. Gerlach is the principal investigator.

replicated this study. He found a correlation of .90 between the 1969 results and his own.

Gerlach and Barron (1974) conducted a study to determine whether or not statements of conditions or statements of criteria might alter individuals' perceptions of the observability of verbs used in complete behavioral objective. For example, would a verb which received a low observability rating when ranked in isolation receive a higher rating if presented within a complete objective with a precise statement of condition and/or statement of criterion. Their results demonstrated that the condition and criterion influence not only the ratings of verbs, but also they influence the ratings of complete objectives more than the verb does. This finding was strengthened by a subsequent study (Barron and Gerlach, 1975), which also revealed that the use of two scales, "most observable to least observable" and "precise to vague," provided more reliable and meaningful data than did the use of two other scales, "specific to general" and "concrete to abstract."

Haygood, Gerlach, and Wigand (1976) extended this line of investigation in a study designed to determine whether or not the direct object in an objective significantly affected the rating. They found that there was no difference between subjects' perceptions of objectives which contained actual direct objects and statements which contained symbolic representations of direct objects (i.e., "x" and "y" were substituted for actual direct objects). However, they did find that varying the actual direct objects did produce significant changes in raters' perceptions of the observability of the objectives.

These studies have led us to conclude that emphasis on any one component of a performance objective is not warranted. This finding is inconsistent with the injunctions and exhortations found in most, if not all, manuals and guidelines for instructional systems design. The studies clearly indicate that empirical evidence is needed to determine how an objective should be constructed in order to maximize the possibility that the reader will perceive it as describing an observable event.

However, all the studies cited above used educators or students in colleges of education as subjects. Generalizing the findings to other populations would be warranted only if a high relationship could be found between the results of the studies cited and the results of a replication using another population. Specifically, it was hypothesized that the findings could be generalized to two types of military personnel: trainees and trainers (including instructional systems designers).

Procedure

The first replication used 30 undergraduate pilot trainees (UPTs) at Williams AFB, Arizona. They were administered an instrument containing 123 verbs which each respondent was asked to rate on a quintile scale from most observable to least observable. This instrument was exactly the same as that used by Deno and Jenkins (1969) and Gerlach (1974), with one addition: the previous instrument contained 99 verbs found in school curricula; the present instrument contained all these plus 23 verbs found in USAF Undergraduate Pilot Training syllabi. A second replication, using the same instrument, was conducted at Luke AFB, where 17 members of four instructional systems development teams served as subjects.

Results

The correlation between the ratings of the UPTs and the results reported by Deno and Jenkins was .90. The correlation between the ratings of the ISD team members and the Deno-Jenkins results was .91. The correlation between the UPT and ISD ratings was .89. The mean rating of the 99 verbs used by Deno and Jenkins was 3.06. UPTs gave these same verbs a mean rating of 2.57, while ISD team members rated them 2.51. The 23 verbs which were randomly selected from ATC training materials were given mean ratings of 2.46 and 2.50 by the UPTs and ISD team members, respectively.

Conclusions

The correlations obtained are sufficiently high to warrant our concluding that, with respect to the content of this study, we are dealing with highly similar if not identical populations. Consequently, we are convinced that the findings of earlier studies concerning the three components of performance objectives, as well as the findings concerning direct objects, are generalizable to the context of Air Force training.

It appears that some of the operational difficulties currently encountered by field ISD teams (especially for complex perceptual motor skill training) would be reduced if existing manuals were revised and new manuals written to incorporate the findings concerning the observability of verbs. Furthermore, the three factor model of objectives (verb, conditions, and criteria) should be expanded to include the direct object.

Tables showing the means and variances of the ratings for each of the 99 verbs for each of the three subject populations will be distributed at the session. Likewise, the tables showing the obtained simple and multiple correlations will be made available.

The Politics of Instructional Innovation in Higher Education: A cross cultural analysis.

Paper for the proceedings of the
Research and Theory Division
Association for Educational Communications and Technology
New Orleans, LA.

March 5th -9th, 1979

John G. Hedberg PhD
Centre for the Study of Higher Education
University of Melbourne
Parkville
Victoria 3052
Australia.

Abstract

The Politics of Instructional Innovation in Higher Education: A Cross Cultural Analysis.

Higher education administrators have supported improvements in teaching by creating special funds to support selected projects. These projects have not been universally successful: some projects have folded after the initial funding ceased, while others never really started. This study was an attempt to test the strength of path models for successful instructional development projects in two different cultures and academic structures. In a series of Australian projects, administrators, instructional developers and project directors were asked to complete the same instrument that had been used in the United States of America. The most important findings have been the completely different factor structures. Where the United States respondents saw the issues as provision of support services, organisational support with good administrative systems, faculty incentives (motivation), and financial resources, the Australian respondents saw the major issues as provision of support and administrative services, status of the project internally and externally (including some similarity to faculty incentives above), positive and clear innovative climate and expertise and skills of the member of faculty to carry out the project.

> "Most of the innovation that occurs could best be described as fiddling. Cummulative assessment is brought in here, open book examining there, a game or two somewhere else, TV, CAI or carrels. These are healthy signs of dissatisfaction and uncertainty, but they tend to relieve the initiators of the need to face the fundamental re-appraisal of objectives, course design, assessment and outcome that experience suggests is both required and possible". (Wilson, 1973).

Innovation and change are fundamental to the process of instructional development, and in recent years administrators in higher education have supported curriculum improvements by funding selected projects. However, these projects have not been universally successful: some projects have folded after the inital funding ceased; some when the project director left for a new position; and some ran into organizational and administrative problems that prevented them from ever starting.

In a review of research on curriculum and instruction, Fullan and Pomfret (1977) established that the major emphases have been on the adoption decision before a change is implemented and monitoring the results of change. Recently, more research has focussed on the process of implementation. Loucks (1978) and Hall (1977) have described their efforts to categorize the level of use of an innovation and thereby predicting future stages of development. Lawrason and Hedberg (1978) have presented a path model to predict instructional development success on the basis of institutional committment and personal expertise of the academic staff. This change in focus provides the instructional development specialist with an ordered array of factors that can be used to successfully manage instructional innovations.

In commenting upon his experience at two universities, Gropper (1977) advises the instructional designer to be sensitive to the bias, fears and misunderstanding on the part of the academic staff and administration unfamiliar with instructional design and change. The current modes of teaching and decision-making are firmly entrenched within th organisational structure. Innovation if it is to occur, must be achieved subtlely and gradually. Support for the project must therefore come from those helping to achieve the change. If the project is supported by the central instructional design, production and evaluation staff, then this will help the initial development of the project considerably, even if funds are not available to support major expenses. In fact, support staff, resources and rewards can be used to ensure the continuation of good projects. Gropper observed that a central facility with the specific expertise in instructional design, evaluation and production should be available centrally for the support of academic staff, and that the administration, academic boards and promotion committees should reward the effective teaching and curriculum design of academic staff participating in instructional design projects. While some administrations rewarded curriculum re-design efforts, by far the majority of projects surveyed in one study (Lawrason,1977) did not have any central instructional development support.

Without support, what then acts as incentives for instructional development? Spitzer (1977) asked academic staff in sixty institutions throughout the United States why they undertook instructional development projects. Spitzer found the indirect organizational incentives (administration committment, competent instructional development staff, etc.,) were the most clearly developed and recognizable success factors. Direct incentives (payments, release time, etc.,), recognition (letters of appreciation, titles, etc.,) and personal satisfaction (achievement, striving for excellence, etc.,) were of lesser importance and demonstrated less consistency in responses appearing to be specific to the individual or the motivation. The evidence of success of most projects as Spitzer pointed out was based upon intuitive rather than research evidence.

Lawrason and Hedberg (1977) identified the interrelations among key instruction design factors. From the instructional design literature they assembled a series of factors that were considered to contribute to project success. These factors, institutional commitment, faculty rewards, instructional design staff expertise, campus AV production services, program evaluation and faculty interest in innovation, were ranked by academic staff, administrators and instructional designers. While administrative commitment to instructional development, primarily through financial support, was extremely important, respondents tended to group factors according to highly individualized concerns, and the factors perceived as contributing to success were related to the respondent's world view, and not to the respondent's academic or administrative rank. The groups of respondents so formed were characterized by their views on: administrative commitment, instructional development policies on curriculum change, and personal satisfaction from undertaking the project. In this study the respondents were asked to reply 'in general' and no attempt was made to relate instructional development success to specific projects.

In a later study, (Lawrason and Hedberg, 1978) respondents were asked to examine a particular instructional design project and comment on its success or failure and the factors that contributed to this outcome. The earlier factors were refined and further sub-divided to produce an instrument that examined four major areas in fifty items; the provision of instructional design and production support services, enthusiasism of the academic staff for change, provision of resources such as money and equipment, the organisational climate that would hinder or facilitate innovations and changes, and items related to student interest and behaviour. One third of the respondents indicated that their project was initiated by a member of the college administration and this fact alone was instrumental in achieving a successful outcome. Under these conditions administrative commitment was active and vigorous. Unfortunately no clear distinction was made between directive and participative generation of the project, thus the role played by the administrator is not always clear. Four measures of project success were collected: an overall estimation of project success, the likelihood of the project continuing if funding ceases, the likelihood of the project continuing if the project director leaves the institution and the effect of the project on student enrollments. Lawrason and Hedberg (1978) found that the likelihood of the project continuing if funding ceased or the director left, were predicted by the institutional and instructional design factors surrounding the project. The other two dependent measures were not predicted by the same factors. The overall estimation of project success was never negative, whereas the two likelihood-of-continuation measures appeared to be less positive and had a greater variance. This may be due to an inherent limitation of this type of study, the respondent was often the project director and initiator which might indicate a degree of ego-involvement with the outcome.

The path models determined for the Lawrason and Hedberg study (figure 1) included both independent variables based on the nature of the project/repondent, and composite intervening variables derived from a cluster analysis of the fifty item questionnaire.

Insert figure 1
about here

The path models were based upon a sample twenty-six instructional development projects undertaken in six universities and colleges throughout the Delaware Valley, Pennsylvania.

Purpose of the current study

This study sought to apply the path models of figure 1 to a similar instructional development environment using a sample drawn from Australian Higher Education Institutions. As with the Lawrason and Hedberg (1978) study, this comparative study is based upon the experience of academic staff who have tried to redesign a course and the factors they felt were important to the success or failure of their instructional project.

Some parallels might be drawn between the two sets of samples.

1. Both samples were predominately undertaking projects to develop their own course of approximately semester length.
2. Most respondents categorized themselves as Departmental Chairman or Tenured Academic Staff. (It is possible that the lower frequency of Untenured Staff responses is due to their changing positions).
3. The most common roles played by respondents were as project director and instructional designer.

Method

The respondents were administrators, project directors, academic staff and educational support staff at two higher education institutions in Victoria, Australia, and six higher education institutions in the Northeastern United States. After each project had been identified through educational research officers and through previous studies, the project director was called on the telephone and asked if he or she and co-workers would reply to the questionnaire. The questionnaire was structured into three sections: Part 1 identified the general nature of the project and the person responding,

1. The title and objective of the project;
2. The size of the project (PSIZE);
3. The academic rank of the respondent (RANK);
4. The part played by the respondent in the project team (PART);
5. The rank of the person who initiated the project (PINIT).

Part II of the survey consisted of a 50-item questionnaire (Appendix A) which measured the perceived importance of a wide range of factors contributing to the outcome of an instructional development project. Respondents were asked to rate each of the 50 items on a scale of 1(low priority) to 7 (high priority). The initial path models had been established with the United States data (Lawrason and Hedberg,1978) and the composite intervening variables derived from the 50 items were:

1. The importance of the relationship between the support staff and the project members (SUPPORT);
2. The financial incentives or intrinsic motivators available to faculty project members (FACINCEN);
3. The financial resources provided by the administration for the funding of the project (RESOURCE); and
4. The importance of organizational facilitation, i.e. the interpersonal relationships between the administration and project members (ORGFAC).

Part III of the survey required respondents first to rate the overall success of their project on the basis of four specific criteria. The final two responses requested subjects to verbally describe the "major factor which facilitated success" and the "major factor which limited the success" of the project. The four specific criteria by which respondents rated the success of their projects were considered as the dependent variables in the study.

They were:

1 A global rating of the overall success of the project (SUCC);
2 The likelihood of project continuation after funds ceased (FUND);
3 The likelihood of project continuation after the original director left (DIR); and
4 The effect of the project upon student enrollments (STUD).

Scores for the composite factor variables (SUPPORT, FACINCEN, RESOURCE, and ORFAC) were used in a standardized form in a regression analysis to determine path co-efficients. On the basis of the Lawrason and Hedberg (1978) study the hypothesized relationships are illustrated in Figure 1. The Australian data was analysized by reference to these previously established path models and then a factor analysis of the intervening variables was attempted to confirm the factor pattern determined from the U.S. data.

Results

The previous study was based upon 38 respondents representing 26 instructional development projects in United States institutions. In this Australian comparison 27 respondents represent 22 projects. While some difficulty was experienced in identifying projects, an 82% response rate was achieved in the two Australian Institutions.

The nature of the surveyed instructional projects was characterized by the size of the project, the respondent's academic rank, the part the respondent played and the rank of the person who initiated the project. (Table 1). No differences were noted in these categories with the exception of the person who initiated the project. Significantly more academic staff initiated instructional projects in the Australian sample, and correspondingly more administrators (deans, departmental chairpersons)initiated projects in the U.S. sample. This may be due to differences in tenure percentages with more tenured staff in the Australian Institutions, or more fluid departmental administrative structure of the American Colleges.

Insert table 1
about here

Comparisions between the two groups were made on the 50-item questionnaire on project success factors. Using simple *t* tests for differences between means, 11 items had *t* values exceeding the probability of .05.

The respondents in the Australian sample considered the following aspects more important to project success than their U.S. colleagues:

11 Provision of administrative rewards (release time, money) for academic staff engaged in instructional projects.
16 Academic staff felt need to change existing course design and teaching strategies.
17 Commitment by academic staff to implement and evaluate the project as designed.
20 Good personal relations between the members of the academic staff.
22 Priority given by support services to full scale funded institutional projects.
26 Efficiency of support services.
27 Positive and supportive attitudes of support service personnel.
28 Good personal relations between academic staff and support staff on the project.
45 Necessity to improve instruction for a specific disadvantaged student population.
46 Large number of students to benefit from the project.

Some observations should be made about these results. The emphasis on academic staff (Q's 11,16,17,20) is likely to be due to the differences in the person who initiated the project. More academic staff were responsible for project initiation in Australian and this would increase the importance of responses directed at academic staff behaviours. The Australian projects required greater media support than the U.S. projects. This would seem to be appropriately relected in items 22,26,27,and 28. One might explain the high rating of the two student benefit items in that the importance of the project to student learning either in terms of quality or quantity has often been used as a criteria for project funding in Australia. Reading the project descriptions requested of each respondent in the questionnaire, it appeared that Australian projects were more closely related to basic instructional problems, while the U.S. projects appeared more to be derived from the respondents research interests. However, the item relating to the importance of the project in generating research findings was not rated differently by the two groups; both rated it low (3.3) compared to other items. Only one item was considered more important by U.S. respondents than Australian respondents, namely

36 Self-support (budget) of project after initial development.

The Australian projects were undertakings that most often had institutional commitment to carry the project into regular funding. This was not always apparent in the U.S. projects.

Factor confirmation

The Australian responses to the 50-items were compared to the factor structure previously determined for the U.S. data (Lawrason and Hedberg, 1978). A similar factor structure was not supported. The only similarity in structures was the continued grouping of items 26,27 and 28, which describe the efficiency of and relationships with the support services. When attempting to confirm the path models of Figure 1 with the previously established factors, no significant paths were found that included the intervening variables: FACINCEN, ORGFAC, SUPPORT, RESOURCE.

In light of this result a factor analysis was undertaken to investigate the structure underlying the Australian responses to the instrument. This produced a solution of four factors which were characterized:

1. The importance of communication and support links between project director, administration and support staff personnel (ORGSUP)
2. The status of the project and the director within and outside the institution (STATUS),
3. The climate or human relationships surrounding the academic staff working on the project, and their ability to be self sufficient (CLIMAT)
4. The instruction design expertise of the project director and the clarity of project goals (EXPERT)

The composition of these intervening variables was confirmed by a cluster analysis using the Veldman H-group program. (Table 2). The simple correlations between the variables used in the final path analysis is given in Table 3.

Insert tables 2& 3
about here

Overall project success

Responses by subjects to the questions about overall sucess of their projects (SUCC, FUND, DIR and STUD) did not differ significantly between the two groups.

Insert table 4
about here

From the direction of the differences in means it would appear that US projects are more dependent on the director for continuation of the project. However they appear to generate more student enrolments than their Australian counterparts.

In writing comments about their projects most Australian directors emphasized the grass roots nature of instructional innovation. While many worked as part of teams, the projects were often self-styled with loosely defined structures. By far the most important factor that helped the project was co-operation from the respondent's colleagues, this was mentioned by two thirds of the respondents. The remaining factors were: funding, positive student reaction, assistance of support staff, clear objectives and time. More serious was the list of limitating factors. One third claimed there was a lack of departmental support and commitment, and that funds were insufficient. Other limitations were delays, not enough people or not enough time. Two projects were hindered by the poor leadership and lack of expertise of the director.

Discussion

Of the projects in this study over half (16) of the respondents were working on single course or small projects. This contrasts markedly with the U.S. sample where the majority of respondents (31) were working on projects larger than a single course. This pattern would seem to reflect different attitudes to instructional development in the two countries. The Australian respondents were primarily working on their own projects, while U.S. respondents often worked on projects initiated by another person.

Many projects generated large amounts of media-based individual study materials. In both samples, the academic staff often undertook the development without an instructional development specialist to assist them. Support personnel were usually in a service role, providing graphics, recording a tape, or consulting on problem areas only. This exclusion of trained instructional development specialists had resulted in some wasteful failures and where specialists were available the lack of good client relationships appeared to alienate the innovative academic staff. An instructional designer could help these projects but only if he or she can gain acceptance as a team member and achieve harmonious relations with the academic staff.

It is interesting to note that only one respondent in the two groups reported a lack of success. The respondent, a chairman of a department, cited the failure of the administration to inform the project director of the degree of funding for the project. The delay, which ran to three years, does seem excessive.

Lawrason and Hedberg (1978) found that of the dependent measures - continuation of funding ceases (FUND) and continuation if the director leaves (DIR) were more important predictors of the successful project. The overall likelihood of a project continuing if funds ceased was .76 for the U.S. sample and .78 for the Australian sample. On the other hand, the likelihood of a project continuing if the director left was .64 for the U.S. sample and .76 for the Australian sample. (The likelihoods ranging from 0-unlikely to 1-certainty). In both sample these two dependent measures were important in the path models and appeared to provide better estimates of the 'real' outcome of a project Although the difference in likelihoods of continuance if the director left was not statistically significant, there is an indication that the U.S. projects are more heavily dependent on their project director for a successful outcome.

Comparision of path models

As previsouly discussed the Australian sample did not conform to the same factor structure nor did it confirm the previously determined path models (Figure 1). The only overlap between the factor structures is the consistent grouping of support services. The remaining factors in the Australian sample emphasize human relations skills, expertise or competence and the standing of the project and director. The factors determined on the basis of the U.S. data emphasized faculty incentives, project resources and administrative commitment, all more external concerns than the more hesitant ego-involved items that were grouped by the Australian sample.

A path model based on the assumption that the four indenpent variables (RANK, PART, PSIZE, PINIT) would predict the four intervening variables (ORGSUP, STATUS, CLIMAT, EXPERT) which in turn would predict the four dependent variables (SUCC, FUND, DIR STUD) produced only limted paths (Figure 2). Unlike the U.S. sample, the overall rating of project success contributed as a major intervening variable in the study. Some similarities are reassuring, such as the importance of the academic rank of the respondent in predicting likelihood of project continuation if the director left. However, the direct RANK to FUND path was not significant in the Australian sample. The path model of Figure 2, does concur with Spitzer (1977) when it emphasises that the climate surrounding the project is an important link in the prediction of project success.

A number of other interesting relationships were indicated by the Australian data (although not included in the final path model): the correlation between organizational support (ORGSUP) and the STATUS of the project; and the joint prediction of instruction development expertise (EXPERT) by the part the respondent played in the project (PART) and the size of the project (PSIZE). This last path pointed out the rather damning relationship - the more important the part the respondent plays in the project and the larger its size, the less important instructional design expertise or competence is rated. Maybe it's simply a fact that large projects suceed in spite of the lack of expertise of the innovative academic staff.

Conclusion

In considering the most essential factor influencing the success or failure of projects, it is interesting to note that respondents place prime attention on the interest, co-operation and skills in the academic staff team members. It is evident that faculty engaged in development projects place the burden of successful implementation upon themselves and their academic colleagues. Over half of those responding in both samples to the question 'major factors contributing to success' mention the initiative or skills of the project director or other academic colleagues. Far fewer list administrative support, support services or even funds.

The comparisions made in this study should raise a note of caution to any researcher extrapolating the findings of one successful innovation to predict outcomes in a different educational system. Even though, close cultural parallels exist in the two samples chosen for this study, differences have been noted in the degree of control over the environment of the project. Australian respondents were more concerned with academic staff behaviour while U.S. respondents were more concerned with institutional reward systems and the provision of resources. The results of this study might be considered a first approximation to the many factors in the institutional environment that affect instructional development. The importance of good interpersonal relations at all levels, mutually defined goals, and the supportive rather than directive role of the development specialist are essential.

References

Diamond, R.M. et al. *Instructional development for individualized learning in higher education*. Englewood Cliffs, N.J.: Prentice Hall, 1975.

Fullan, M. & Pomfret, A. Research on curriculum and instruction implementation. *Review of Educational Research*, 1977, *47*, 335-393.

Gropper, G.L. On gaining acceptance for instructional design in University settings. *Educational Technology*, 1977, *17*(12), 7-13.

Hall, G.E. and Loucks, S.F. A developmental model for determining whether the treatment is actually implemented. *American Educational Research Journal*. 1977, *14*, 263-276.

Lawrason, R.E. & Hedberg, J.G. A systems model for instructional development success factors? *Educational Technology*, 1977, *17*(4), 15-20.

Lawrason, R.E. Politics and diffusion strategies for instructional development in higher education. Paper presented at the Convention of Association of Educational Communications and Technology, Miami Beach, April 1977. Resources in Education, 1977, *12*(11). ERIC Document ED 140 769.

Lawrason, R.E. & Hedberg, J.G. Predicting successful instructional development projects in higher Education. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 27th, 1978.

Loucks, S.F. Conceptualizing and Measuring program implementation: A variable useful for planned change and evaluation. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 27th, 1978.

Spitzer, D.R. Incentives in instructional development. *Educational Technology* 1977, *17*(8), 17-21.

Wilson, R.K. Submission to the committee of inquiry on the Open University. Canberra: Australian University Commission, 1973.

Acknowledgements

Dr. Robin Lawrason of Temple University provided the initial impetus for this study and worked to collect data on the United States sample.

Mrs. Claire White at the University of Melbourne assisted with the path analysis.

Appendix A

Part II

The following factors can contribute to the success of an educational development project within an institution Please rate the importance of each factor as you believe it contributed to the success of the educational project you have identified. Circle the number that most closely represents the importance

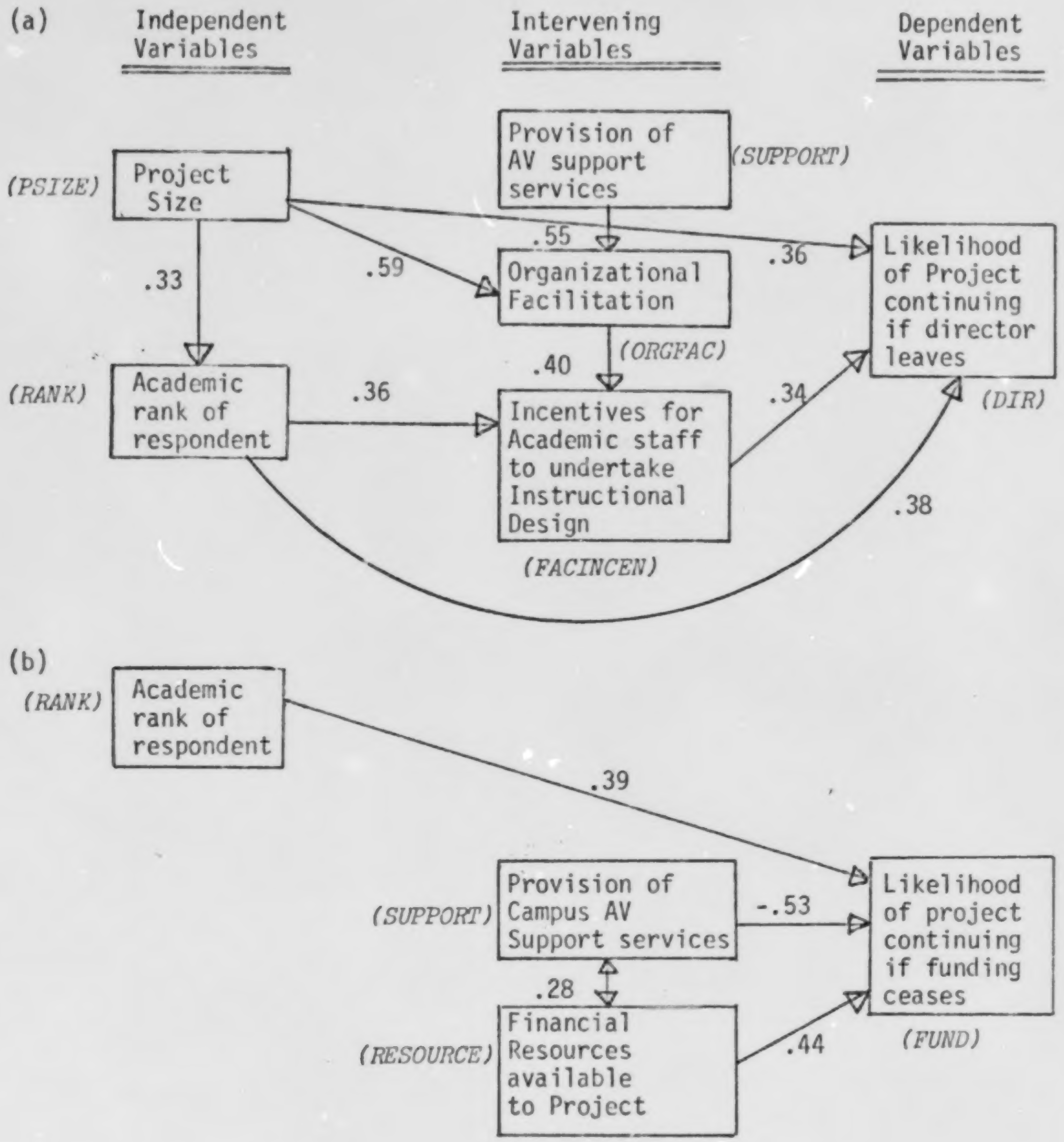
For example:

	Rapid provision of typed manuscripts	Unimportant	1 2 [✓ 3] 4 5 6 7	Important
1	Clearly defined organizational structure for decision making and implementation	Unimportant	1 2 3 4 5 6 7	Important
2	Small numbers of people involved in the decision making associated with the project	Unimportant	1 2 3 4 5 6 7	Important
3	Lack of competitiveness and possessiveness within existing academic departments	Unimportant	1 2 3 4 5 6 7	Important
4	Personal support of top administrators	Unimportant	1 2 3 4 5 6 7	Important
5	Provision of instructional project 'start-up' funds from operating budgets	Unimportant	1 2 3 4 5 6 7	Important
6	Continued budgetary support of administrators	Unimportant	1 2 3 4 5 6 7	Important
7	Administration concern for student enrolments	Unimportant	1 2 3 4 5 6 7	Important
8	Flexibility of administrators to integrate change into the administrative system	Unimportant	1 2 3 4 5 6 7	Important
9	Positive leadership skills of administrators	Unimportant	1 2 3 4 5 6 7	Important
10	Initiation of projects by administrators	Unimportant	1 2 3 4 5 6 7	Important
11	Provision of administrative rewards (e.g. release time, money) for academic staff engaged in instructional projects	Unimportant	1 2 3 4 5 6 7	Important
12	Direct communication between administration and academic staff	Unimportant	1 2 3 4 5 6 7	Important
13	Large number of people involved in the decision making associated with the project	Unimportant	1 2 3 4 5 6 7	Important
14	Initiation of projects by individual members of the academic staff	Unimportant	1 2 3 4 5 6 7	Important
15	Promotional considerations given to academic staff for work on projects	Unimportant	1 2 3 4 5 6 7	Important
16	Academic staff felt need to change existing course design or teaching strategies	Unimportant	1 2 3 4 5 6 7	Important
17	Commitment by academic staff to implement and evaluate project as designed	Unimportant	1 2 3 4 5 6 7	Important
18	Student positive feedback to academic staff on course design or improvement	Unimportant	1 2 3 4 5 6 7	Important
19	Openness and encouragement from department chairperson for instructional improvement	Unimportant	1 2 3 4 5 6 7	Important
20	Good personal relations between members of academic staff	Unimportant	1 2 3 4 5 6 7	Important
21	Experience of academic staff in course design, development and evaluation	Unimportant	1 2 3 4 5 6 7	Important
22	Priority given by support services to full scale funded institutional projects	Unimportant	1 2 3 4 5 6 7	Important

23	Frequent communication between administration, academic staff, consultants, and support staff on progress of project	Unimportant	1 2 3 4 5 6 7	Important
24	Project planning actively involved input from academic staff, administration, instructional design consultants and other support services	Unimportant	1 2 3 4 5 6 7	Important
25	High quality of materials produced for project	Unimportant	1 2 3 4 5 6 7	Important
26	Efficiency of support services	Unimportant	1 2 3 4 5 6 7	Important
27	Positive and supportive attitudes of support service personnel	Unimportant	1 2 3 4 5 6 7	Important
28	Good personal relations between academic staff and support staff on the project	Unimportant	1 2 3 4 5 6 7	Important
29	Availability of professional consultants to assist with educational principles related to the design and evaluation of projects	Unimportant	1 2 3 4 5 6 7	Important
30	Diversity of technical support services, e.g. number of alternative media available, CAI, TV, audio, print, graphics, clerical, etc.	Unimportant	1 2 3 4 5 6 7	Important
31	Responsible position of project director in relation to institutional hierarchy	Unimportant	1 2 3 4 5 6 7	Important
32	Interpersonal skills of project director, e.g. accessibility, political acumen, team leadership, etc.	Unimportant	1 2 3 4 5 6 7	Important
33	Clarity of project objectives	Unimportant	1 2 3 4 5 6 7	Important
34	Initiation of projects by project director	Unimportant	1 2 3 4 5 6 7	Important
35	Attractiveness of teaching/learning techniques to potential students	Unimportant	1 2 3 4 5 6 7	Important
36	Self-support (budget) of project after initial development	Unimportant	1 2 3 4 5 6 7	Important
37	Attractiveness of subject matter to potential students	Unimportant	1 2 3 4 5 6 7	Important
38	Several academic staff members actively working on the development of the project	Unimportant	1 2 3 4 5 6 7	Important
39	One or two academic staff members actively working on the development of the project	Unimportant	1 2 3 4 5 6 7	Important
40	Potential of the project to generate educational research findings	Unimportant	1 2 3 4 5 6 7	Important
41	A number of publications generated by the project	Unimportant	1 2 3 4 5 6 7	Important
42	Availability of funds for project from external funding agencies	Unimportant	1 2 3 4 5 6 7	Important
43	Marketability of end product outside institution	Unimportant	1 2 3 4 5 6 7	Important
44	Instructional design skills of project director	Unimportant	1 2 3 4 5 6 7	Important
45	Necessity to improve instruction for a specific disadvantaged student population	Unimportant	1 2 3 4 5 6 7	Important
46	Large number of students to benefit from project	Unimportant	1 2 3 4 5 6 7	Important
47	Project reflects response to current social needs	Unimportant	1 2 3 4 5 6 7	Important
48	Formal evaluation of the instructional effectiveness of the project	Unimportant	1 2 3 4 5 6 7	Important
49	Good personal relations between project director and administration	Unimportant	1 2 3 4 5 6 7	Important
50	Prestigious nature of the project throughout the institution	Unimportant	1 2 3 4 5 6 7	Important

©1978 J.G. Hedberg & R.E. Lawrason

Figure 1 Path Diagrams for Successful Instructional Projects (Lawrason & Hedberg, 1978)


(a)
Independent Variables
Intervening Variables
Dependent Variables
(PSIZE)
Project Size
Provision of AV support services
(SUPPORT)
.55
.59
.33
.36
Organizational Facilitation
Likelihood of Project continuing if director leaves
(ORGFAC)
.40
(RANK)
Academic rank of respondent
.36
Incentives for Academic staff to undertake Instructional Design
.34
(DIR)
.38
(FACINCEN)
(b)
(RANK)
Academic rank of respondent
.39
Provision of Campus AV Support services
(SUPPORT)
-.53
Likelihood of project continuing if funding ceases
.28
Financial Resources available to Project
(RESOURCE)
.44
(FUND)

1 The words in italics and in parentheses are shorthand forms used in the statistical program.

Figure 2 Path diagrams for successful instructional projects based upon an Australian sample.

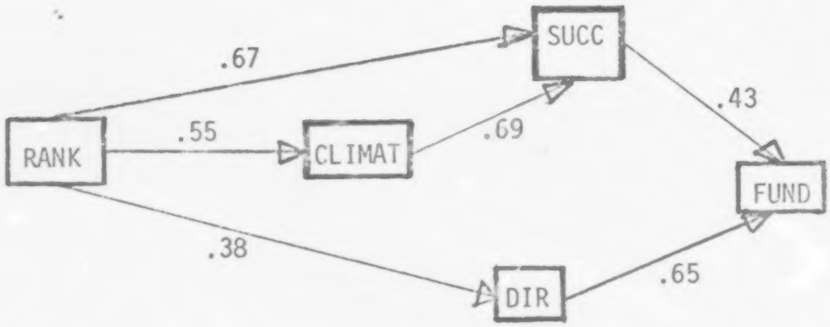

SUCC
.67
.43
RANK
.55
CLIMAT
.69
FUND
.38
.65
DIR

Table 1 The Nature of the Projects

Variable	Category	U.S. Frequencies N= 38 *	Australian frequencies N= 27 *
PSIZE	1 Smaller than a course	2	6
The size of the project	2 One course	5	7
	3 More than one course	31	10
	4 General project related to teaching	0	3
RANK	1 Senior administrator	7	3
The academic rank of the respondent	2 Departmental Chair	4	4
	3 Tenured faculty	21	17
	4 Non-tneured faculty	5	2
	5 Support staff	1	1
PART	1 Project director	22	18
Part respondent played in the project team	2 Instructional designer	3	5
	3 General support	11	3
	4 Not a member	2	1
PINIT	1 Administrator	22	3
Rank of person who initiated the project	2 Faculty	15	21
	3 Support staff	1	2

* If the totals are less than 38 or 27 respectively, the remaining responses are missing.

Table 2 Composition of Factors contributing to instructional innovation success. (Based upon Australian Sample. N=27)

ORGSUP	Items 5, 12, 23, 26, 27, 28, 49
STATUS	Items 10, 15, 24, 30, 32, 39, 40, 41, 43, 50
CLIMAT	Items 3, 4, 19, 20, 35, 37, 42*
EXPERT	Items 17, 21, 33, 34, 44

---

*Item 42 is reverse scored.

Table 4 Comparison of Dependent Measures

Variable	Mean U.S. data (N= 38)	Mean Australian data (N= 27)	*t*-value	p
SUCC	5.8	5.4	-1.30	.19
FUND	5.3	5.5	0.34	.74
DIR	4.5	5.4	1.39	.17
STUD*	5.3	4.8	-1.31	.20

* This question was omited by many respondents - samples were 28 and 20 respectively for this question only

Table 3 Simple correlations between variables in the Australian sample

	PSIZE	RANK	PART	PINIT	ORGSUP	STATUS	CLIM	EXPERT	SUCC	FUND	DIR
RANK	.08										
PART	.25	-.10									
PINIT	.03	-.05	.07								
ORGSUP	.11	.07	-.24	-.01							
STATUS	.34	.27	-.27	.04	.40						
CLIM	-.05	-.54	.09	.11	.15	.07					
EXPERT	-.30	-.12	.31	-.02	.05	.17	.38				
SUCC	-.25	.29	.05	.04	.12	.26	.32	.37			
FUND	-.12	.13	.16	.06	-.21	-.07	-.10	.09	.42		
DIR	.02	-.38	.15	-.18	-.10	-.15	.00	-.02	.00	.65	
STUD	-.07	-.18	.13	-.14	.04	-.15	.29	.10	.28	-.17	.08

EFFECT ON INTENDED AND INCIDENTAL LEARNING FROM THE USE OF LEARNING OBJECTIVES WITH AN AUDIOVISUAL PRESENTATION

A paper presented to the Research and Theory
Division of the 1979 AECT Convention by
Robert Main, Associate Professor at the Center
for Information and Communication Studies,
California State University, Chico, California.

ABSTRACT

This paper reports a controlled field experiment conducted to determine the effect and interaction of five independent varibles with an audiovisual slide-tape program: presence of learning objectives, location of learning objectives, type of knowledge, sex of learner, and retention of learning. Participants were university students in a general studies course. Test scores measuring intended and incidental knowledge were the dependent variables. The major finding supported Ausubel's subsumption theory in that use of learning objectives as an advance organizer presented before the slide-tape programs significantly increased intended learning. No significant loss was found for incidental learning. Indeed, from the direction of the data, the tendency was for the learning objectives to improve rather than lessen incidental learning. The use of the learning objectives before the presentation had no significant effects on the affective variables measured and did not detract from the students' evaluation of the program. To the contrary, the trend of the data indicated students receiving learning objectives prior to the presentation liked the program more than when no learning objectives were used.

Introduction and Background

The nonprint media are becoming increasingly important as educational tools for American schools. The increase in availability of educational media materials for school use has been dramatic. Twenty-five years ago, perhaps 25,000 nonprint instructional programs might have been available to the elementary and secondary schools of this country.[1] Today, the number of titles is nearing 500,000.[2] Because instructional media of many kinds are important in structuring what goes on in classrooms and what students do to learn, consideration of ways to improve their use is important.

It would be beneficial for teachers to have practical methods for influencing the efficiency and effectiveness of audiovisual materials selected for curriculum support. There is little in the literature to guide the teacher in the application of instructional theory to the nonprint media. The empirical evidence regarding learning objectives, advance organizers, and mathemagenic activities to facilitate learning is largely concerned with print materials.

Printed curriculum materials allow the learner a wide latitude in use--browsing, skipping, review, repetition. Through selection, pacing, and sequencing, the learner may individualize the material to his particular needs or preference. Audiovisual materials, on the other hand, are commonly fixed pace, fixed sequence presentations with content selected and organized for use with as wide an audience as possible.

The purpose of this study was to investigate the effects of the use of learning objectives with an audiovisual presentation on the intended (objective relevant) and incidental (non-objective relevant) learning outcomes. This study was designed to provide evidence regarding the facilitative effects of using learning objectives as an organizer with a fixed-pace, fixed-format, non-print medium.

Theoretical Foundations

The primary theoretical base for this study is anchored in the cognitive construct instructional theories and the concept of external organizers to facilitate learning as developed by Ausubel.[3] The use of learning objectives as organizing tools is derived from Gagne's principles,[4] and the concept of modifying learning activity by differential placement of the organizing device is developed from Rothkopf's[5] hypotheses of mathemagenic behavior.

In opposition to the more behaviorally oriented theories dominated by Skinner,[6] the cognitive construct instructional theorists attribute a greater degree of autonomy and initiative to the learner. Typically, they emphasize the student's internal representations or cognitive constructs as the really important concern in education.

Advance Organizers

Of particular interest to this study is the emphasis on organization as a critical component of the learning process as espoused by David P. Ausubel and others.[7] Ausubel has articulated these ideas into a theory of meaningful verbal learning which he calls the subsumption theory.[8]

Briefly, the theory holds that meaningful verbal learning can be incorporated into the learner's cognitive structure and there be retained more accessible, and for longer periods of time, if the cognitive structure initially contains discernible, relevant, and related bits of information (subsumers) which the learner can associate with new learning. With the passage of time, the incorporated learning will undergo successive reformations (obliterative subsumption) as the learner attempts to condense and hierarchically arrange added learning.

Ausubel[9] has supported his theory with experimental evidence which indicates that (1) those students who are known to have relevant subsumers will remember more of the subsequent learning for a longer period of time compared to those who do not; and (2) low achievers exhibit greater increases in achievement with the aid of subsumers than high achievers. Ausubel attributes this effect to the apparent lack of organization in the cognitive structure of the low achievers. Externally imposed "advance organizers" provide the necessary organization in low achievers but are of little aid to high achievers who impose their own organization on unstructured material.

According to Ausubel and Robinson, optimal utilization of the principles of progressive differentiation and integrative reconciliation can be achieved by the "...supplementary availability of a hierarchical series of advance 'organizers.'"[10] True organizers, in Ausubel and Robinson's view, are not to be confused with ordinary introductory overviews which are typically written at the same level of abstraction, generality, and inclusiveness as the learning material and achieve their effect largely through repetition, condensation, selective emphasis on central concepts, and pre-familiarization of the learner with certain key words. They considered summaries as comparable to overviews, but probably less effective because their influence on cognitive structure is retroactive rather than proactive relative to the learning task. They felt overviews and summaries "insofar as they imply to some learners that the material they do not include is relatively superfluous . . . may promote neglect of and failure to study or review much significant subject matter."[11]

Ausubel and Robinson also felt advance organizers had "certain inherent advantages" over various kinds of intramaterial organization (adjunct organizing aids within the body of the material).[12] They reasoned that existing subsumers in cognitive structure lacked both particularized relevance for the new material (since the learner could not possibly anticipate its precise nature) as well as the sophisticated knowledge of subject matter and pedagogy available to expert programers.

Behavioral Objectives

The avenue to be explored by this study was the extent that learning objectives could be used as organizing devices for audiovisual presentations. Gagne is a principal proponent of the use of behavioral objectives as facilitators of learning. He states:

> Defining objectives for the student is an inadequately exploited educational technique . . . . (U)nless students know what the objectives are, they are likely to resort to memorization and mechanical completion of the exercises in textbooks or workbooks rather than carrying out relevant sorts of learning activities. When one tells the student what he is expected to do after he learns, he is not "giving him the answer." Rather it is providing him with a goal which he himself can use to organize his own learning activities.[13]

The use of objectives as organizational devices is also supported by Garrison and Magoon who state, "Organization is the process which converts the raw materials of learning into a constructive approach toward meeting an educational objective."[14] They emphasize the importance of objectives as a learning aid by the students in interacting with the instructional material by stating: "Learning experiences should correlate closely with the established objectives of which the learner has been fully apprised."[15]

Ellis also links objectives and organization as important conditions to learning from the student's viewpoint.

> A major feature of good study habits is that you understand the principle objectives or goals of whatever it is you are studying . . . . (K)eep in mind that you are trying to get a good "mental roadmap" of the material before you study it in greater detail. . . . If structure is not immediately evident in the material, then you must seek structure or create some kind of useful system for organizing the materials. This is simply to say that you must impose some kind of organization upon the materials . . . .[16]

Mathemagenic Behavior

Another theoretical area relevant to this study is the concept of mathemagenic behavior developed by Rothkopf and his associates. Rothkopf asserts that the effect of an instructional document can be affected to some degree by controlling the behaviors of the learner in relation to the document. Since these behaviors affect learning, Rothkopf has called them mathemagenic behaviors, a word of Greek origin which means behaviors "which give birth to learning."[17] He postulates that there is a variety of activities in which the instructor or learner can engage, which can modify, and to a degree control, mathemagenic behaviors.

Most of Rothkopf's research involves the use of questions inserted in written materials as a means of modifying mathemagenic behavior to increase learning. There have been a number of studies by Rothkopf and others to support his hypothesis. Several variables have been examined in the experiments but they all have one variable in common--question location. In general, Rothkopf's studies indicated that grouped pre-questions and adjunct pre- and post-questions resulted in greater learning of question-specific relevant knowledge than from either a no-question or grouped post-question treatment. He also found more learning of non-relevant question information by the adjunct post-question treatment only. The frequency of adjunct post-questions also tended to aid learning.[18]

Rothkopf suggests that pre-questions serve a cueing function of informing and directing the learner regarding important content. Post-questions do not affect selective attendance to the material but rather force a mental review of all material (both intended and incidental knowledge) in order to answer the question.[19]

Intended and Incidental Learning

The facilitative effects of using learning objectives as directions to learn relevant prose text material has been well documented.[20] The relative effects of objectives with respect to incidental (non-objective relevant) learning are less clear. Several studies have reported greater incidental learning for groups receiving objectives (or questions) than for non-objective control groups.[21] Other investigations have shown a decrease in learning performance when instructional objectives are presented before the learning materials. Duchastel and Brown,[22] for example, found that subjects who were given objectives performed somewhat more poorly on incidental material than a control group who received no objectives. The reason for the difference between the results of these experiments is not clear. Kaplan and Simmons[23] tested several hypotheses concerning the reasons for the observed increases in learning found in these studies. Specifically, they investigated the effects of the location of the instructional objectives with respect to the learning material (before vs. after). Their principal finding was that performance on information relevant to an objective was relatively high whether the objectives were presented before or after the text. Performance on incidental material was greater for objectives located after the text than before the text. These results are generally supported in other studies by Bruning,[24] Frase,[25] and Rothkopf[26] which show that overall retention tends to be greater when questions are located after rather than before text segments.

The general conclusion is that objectives or questions presented after the learning materials cannot be used as orienting stimuli which permit the learner to selectively attend to specific information in the text.

Virtually all of the research concerning behavioral objectives as organizing devices has been concerned with printed material. Lumsdaine, May and Hadsell[27] studied the effects on learning of supplementary instruction presented before or after a film. While they did not use "learning objectives" they did employ a variety of techniques to direct attention to specific information in the film content. Although intended learning did increase, the effects were small and the results inconclusive, particularly with regard to any effect to incidental learning. Dayton[28] conducted a recent investigation upon the effects of inserted questions in a slide-tape presentation. He found a significant increase in the amount of question-relevant learning but the effects on incidental learning were ambiguous.

In light of these studies, the following questions may be asked:

1. Does the insertion of learning objectives in slide-tape presentations result in greater learning than would occur as a result of a slide-tape presentation only?
2. What effect does the location of the learning objectives in respect to the presentation have on student learning from slide-tape materials?
3. What are the differential effects of the placement of learning objectives upon the intended and incidental learning by the student from slide-tape presentations, i.e., is the effect generally facilitating (results in greater learning of all content) or is it due only to review (results in greater learning of material relevant only to the learning objectives)?

Answers to these questions should provide useful information for people involved in the use of media as an instructional tool. Most of the audiovisual presentations available have not been designed with specific terminal behavior patterns in mind. Frequently the materials used in instruction were not even intended for educational purposes. It would be of value to teachers, to producers of audiovisual programs, and to curriculum developers to know the effects the introduction of learning objectives to the program will have on facilitating the intended knowledge acquisition. The addition of learning objectives would add very little to initial production costs and could be added for minimal costs to existing inventories of slide and video programs. It would be more costly to insert learning objectives in motion picture prints, but supplemental projection of objectives by 35 mm projectors would be practical when using films.

This study may be of equal interest in defining the effect of the insertion of learning objectives on the acquisition of incidental knowledge. There has been a great rush to increase the use of behavioral objectives and their efficacy in improving learning outcomes is well documented. Less attention has been paid to their potential in terms of their effects on incidental learning. The results of this study should serve not only as a guide to benefits of inserting learning objectives in audiovisual presentations, but also as an indication of possible side effects of loss in incidental knowledge. This caution may be of significance in certain situations where the emphasis is to expand abilities to synthesize material rather than focus on information transfer. In addition, the answers to these questions would have implications in determining the validity of the propositions by Ausubel and Gagne and by expanding the implications to include the use of learning objectives as organizers for audiovisual materials.

Study Design

A controlled field experiment was conducted to measure the intended and incidental learning outcomes resulting from the use of learning objectives with a selected audiovisual presentation.

The experiment was designed to determine the effect and interaction of five independent variables with an audiovisual slide-tape program: presence of learning objectives, location of learning objectives, type of knowledge, sex of learner, and retention of learning. A 4x2x2 factorial design replicated for retention of learning was selected for the analysis. Student learning of intended knowledge and incidental knowledge, as measured by paper and pencil tests, were the dependent variables. The design model is at Figure 1.

A 20 minute commercially produced educational slide-tape presentation concerning advances in mass media and communications was used as the stimulus.[29] The participants were students enrolled in an introductory mass communication course at California State University, Chico, in the spring semester, 1977. The course was required for all communication majors and was, in addition, a service course for the University's general studies program. The course content was heavily mediated with almost every class meeting using a film, videotape, audiotape, or multimedia presentation to support the instruction. Students in the class were divided into three treatment groups and one control group by a random sampling procedure stratified for sex.

The treatment factor was the presence or absence of learning objectives. Learning objectives were used with the slide-tape presentation for experimental groups 1, 2, and 3. No learning objectives were available for experimental group 4 which constituted a control group.

The learning objectives for the course content contained in the slide-tape treatment material were established by interviewing the professor of the

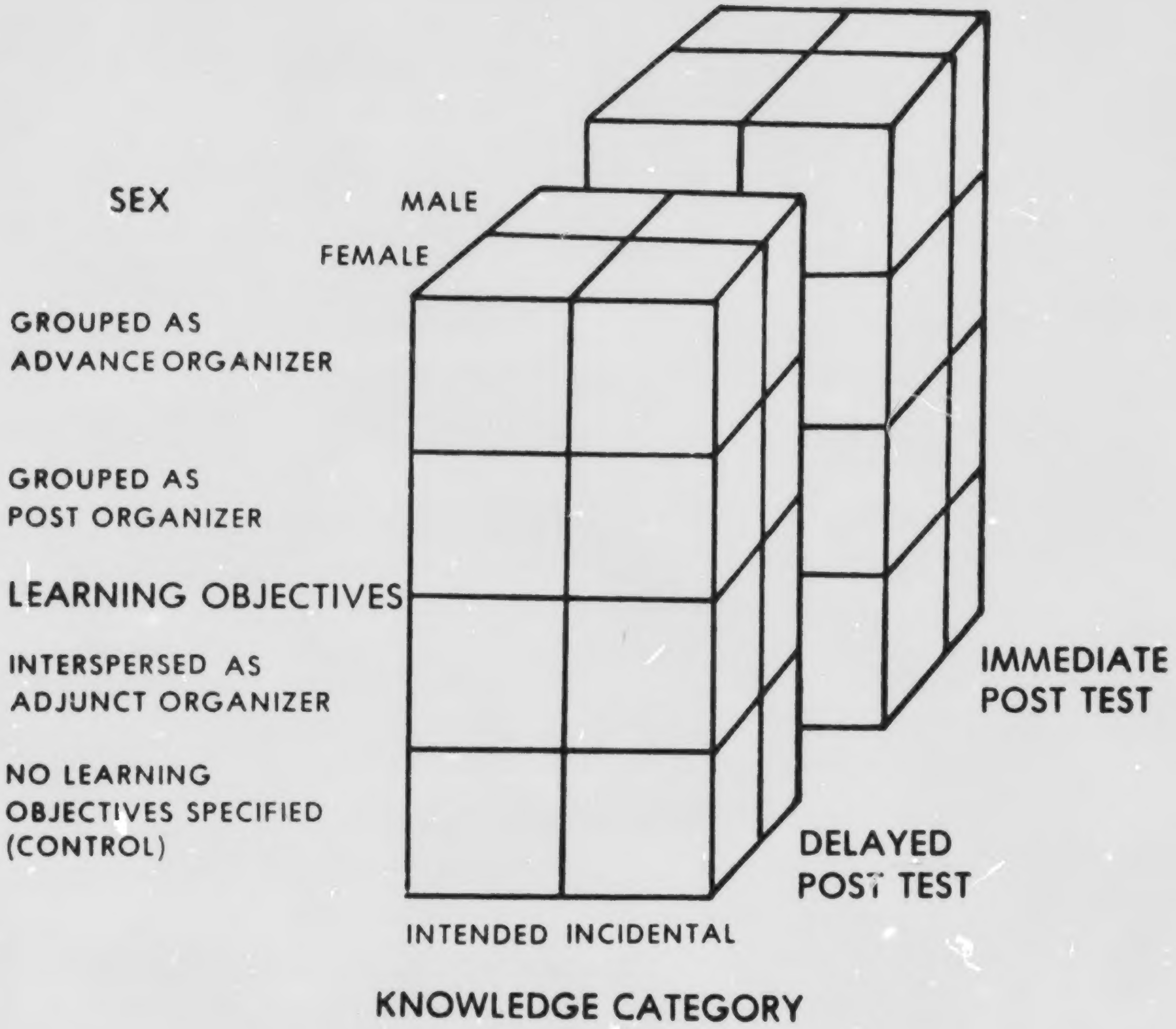

SEX
MALE
FEMALE
GROUPED AS ADVANCE ORGANIZER
GROUPED AS POST ORGANIZER
LEARNING OBJECTIVES
INTERSPERSED AS ADJUNCT ORGANIZER
NO LEARNING OBJECTIVES SPECIFIED (CONTROL)
IMMEDIATE POST TEST
DELAYED POST TEST
INTENDED
INCIDENTAL
KNOWLEDGE CATEGORY

Figure 1

Experiment Model

The experiment uses a two-way analysis of variance including treatment (presence and location of learning objectives) and sex (male-female). The delayed intended and incidental knowledge posttest scores are subjected to identical analysis of variance allowing comparison of data. The immediate posttest scores are treated as a separate but identical experiment.

course, taking her educational aims and specifying the terminal behavior requirements. The initial objectives were then submitted to five professors who had previously taught the introductory mass communication course and were thus familiar with the educational aims of the particular block of instruction, but were not familiar with the slide-tape program. Each professor was asked to rank-order the learning objectives, modify as appropriate, and add or delete objectives if necessary. In addition, each professor was asked to furnish three questions designed to measure by paper-pencil method the intended behavior specified by each objective. From this modified list of learning objectives, four receiving consensual judgment as being appropriate were selected. Each learning objective was then reproduced on a 35 mm slide.

The location of treatment factor was the application of the learning objectives inserted as a group before the presentation (advance organizer), inserted as a group after the presentation (post organizer), and interspersed within the presentation (adjunct organizers). The audio track for the presentation was not changed except to add 10 seconds of silence for each behavioral objective slide.

The type of knowledge factor constituted the cognitive level of information relevant to the learning objectives (intended knowledge) and the cognitive level of information non-relevant to the learning objectives (incidental knowledge).

The intended and incidental knowledge content of the slide-tape presentation was determined by a content analysis. The program was presented to a graduate class in instructional theory and design. Copies of the script and the learning objectives were furnished each student, who was then asked to list the major points of information contained in the presentation that did not directly contribute to the learning objectives. This body of information was designated as incidental knowledge and test questions were developed to measure the cognitive level of this incidental knowledge. All other information was considered intended or objective-relevant knowledge.

The instrument used to measure knowledge in the experiment was a pencil and paper test with three parts. Part one measured intended learning and consisted of 10 items provided by the learning objectives development panel as discrete measurements of intended learning behavior.

Part two of the instrument measured incidental learning and consisted of 12 items developed during the content analysis by the instructional theory and design class.

Part three of the instrument measured affective variables concerning the subject matter and the presentation. It consisted of 10 items each with a series of statements arranged in a Likert type semantic differential scale ranging in value from one through five. The variables measured were: (1) optimism concerning the contribution to society of communication technology, (2) the importance of communication media to society, (3) relevance of material to the class, (4) bias or objectivity of the presentation, (5) interest in the subject, and (6) evaluation of the presentation. The evaluation of the presentation was further defined by five categories:

1. Overall evaluation
2. Organization
3. Length
4. Quality of visuals
5. Format

The test items for measuring the affective domain were developed by the researcher and were not subjected to a validity or reliability test.

A pilot study to establish the validity of the test instruments was conducted by presenting the audiovisual presentation to 33 students enrolled in an undergraduate course in broadcast management. No behavioral objectives were given to the students. The test was administered immediately following the presentation.

The results of the pilot test indicated the internal consistency of the test instrument was acceptable with a Kuder-Richardson reliability of .75. This was considered satisfactory for the experiment's purpose.[30]

No analysis was made of the affective measurement items other than inspection to determine that variance was present. The learning retention factor was the cognitive level of intentional and incidental information measured immediately following the presentation and measured again after a 14 day time lapse.

The experiment was conducted on May 3, 1977 during the normally scheduled class meeting time. Viewing and listening factors were carefully controlled for each group. The subjects were directed by name to four separate and identical classrooms equipped with identical 35 mm Kodak slide projectors synchronized by identical Wollensak audiotape playback machines. Table 1 is the schedule of presentation treatments. The subjects observed the slide-tape presentations without additional instruction. Immediately following the presentation, a graduate assistant administered the examination. All questions were contained in a test booklet and the students recorded their responses on a standard machine graded answer sheet. This was the normal examination procedure employed for the class and consequently the students were familiar with the procedure.

Table 1

Presentation of Treatments

Group	Schedule
1	Learning objectives presented as a group in advance of slide-tape presentation
2	Learning objectives presented as a group at conclusion of slide-tape presentation
3	Learning objectives presented individually throughout presentation immediately preceding relevant portion of slide-tape presentation
4	Slide-tape presentation only; no learning objectives provided

In the follow-up test conducted on May 17, 1977, the subjects were given identical test booklets in their regular classroom, answered the questions, and turned in the booklets at the end of ten minutes. Part III (affective measurement) was deleted from the follow-up test booklets.

To avoid as much as possible the effects of learning from the test, the test booklets were collected after each post-test, and the results of the immediate post-test were not announced to the participants.

The Results

The raw test data collected on each student participating in the experiment consisted of 108 matched immediate and delayed post-test scores, equally divided into three treatment groups and a control group. The number and source of raw test data scores collected on each student and used for the analysis are listed in Table 2.

Additional data items on each participant were collected from student records maintained by the class instructor. These were sex, class level and grade point average (GPA) for the course. A chi-square statistic was computed

Table 2

Data Scores Available for Analysis

Data Item	Number and Source of Scores	
	Immediate Post-test	Delayed Post-test
Intended learning	108	108
Treatment 1	27	27
Treatment 2	27	27
Treatment 3	27	27
Control (no treatment)	27	27
Incidental learning	108	108
Treatment 1	27	27
Treatment 2	27	27
Treatment 3	27	27
Control (no treatment)	27	27
Affective domain		
Optimism	108	
Importance	108	
Bias	108	
Evaluation of presentation		
Overall	108	
Relevance	108	
Organization	108	
Length	108	
Visual quality	108	
Format	108	
Interest	108	

to test the "goodness of fit" of the sample distributions across each of these characteristics. The results indicated the sample groups were relatively homogeneous for the characteristics observed with no significant differences existing between distributions across sex, class level or grade point average.

The mean score of each treatment group for the intended and incidental knowledge measurements are displayed in Table 3. A graphic presentation of the dependent variables in percentage figures is shown in Figure 2.

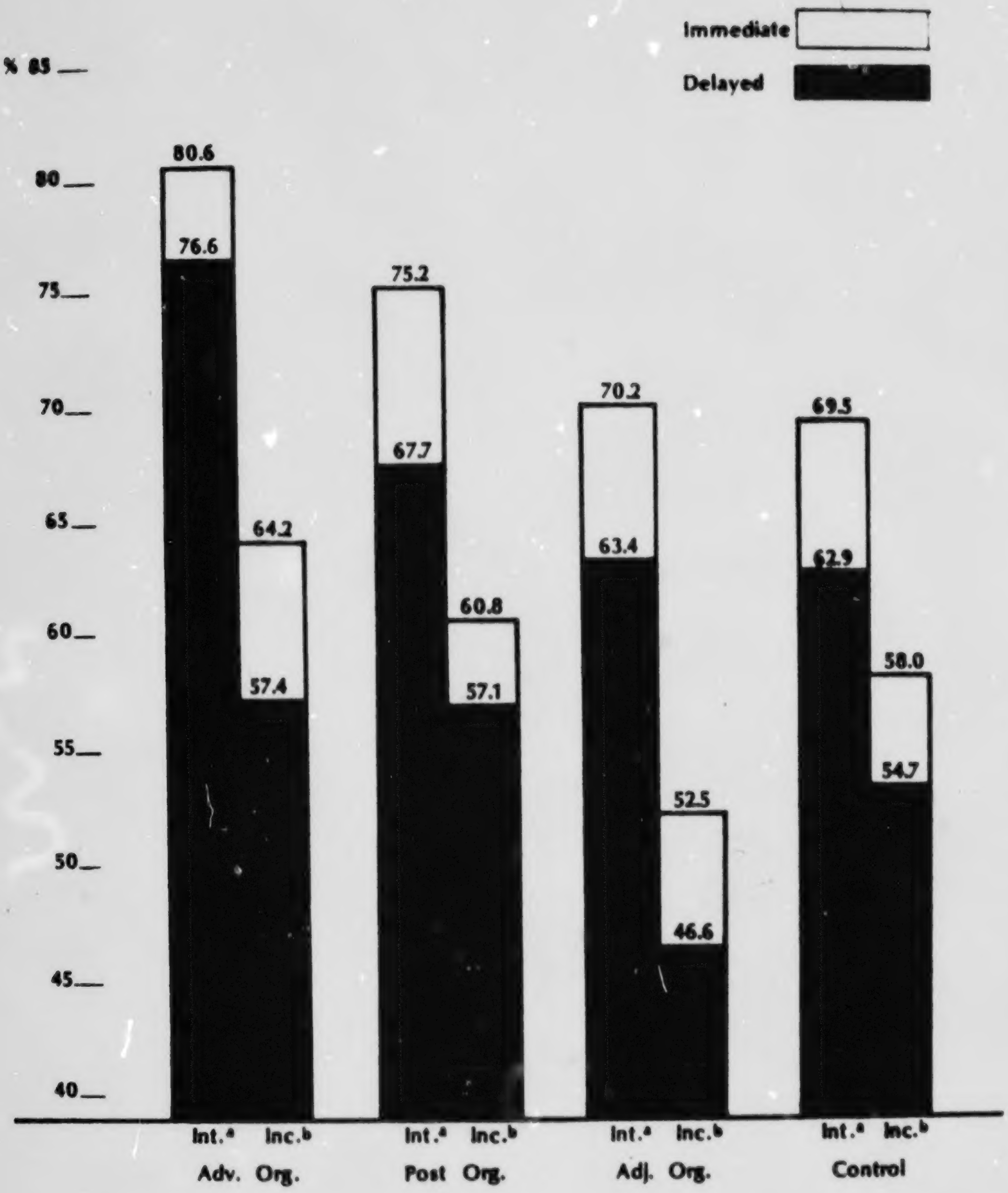


Figure 2

Pattern of Dependent Variable Scores by
Knowledge Type and Experimental Group

---

[a] Intended knowledge percentage score.

[b] Incidental knowledge percentage score.

Table 3

Mean Scores of Test Responses

		Treatment Group			
Knowledge Type		Advance Org.	Post Org.	Adjunct Org.	Control
Delayed Intended	Mean	7.66	6.77	6.34	6.29
	S.D.	2.4	2.4	2.4	2.5
Immediate Intended	Mean	8.06	7.52	7.02	6.95
	S.D.	2.3	2.3	2.3	2.2
Delayed Incidental	Mean	6.89	6.85	5.59	6.56
	S.D.	1.9	2.3	2.2	2.0
Immediate Incidental	Mean	7.70	7.30	6.30	6.96
	S.D.	2.1	2.3	2.3	1.8

The Hypotheses Tests

There were three research hypotheses applied to the study. They were that under the conditions of the experiment:

1. There is no significant difference in cognitive learning outcomes of objective relevant (intended) knowledge in an audiovisual presentation as a result of the insertion of learning objectives.
2. There is no significant difference in cognitive learning outcomes of incidental knowledge in an audiovisual presentation as a result of the insertion of learning objectives.
3. Inserting learning objectives at the beginning, at the end, or interspersed during an audiovisual presentation has no significant difference in cognitive learning outcomes of either intended or incidental knowledge.

In the analysis, incidental and intended learning were considered separately using a two-way analysis of variance with treatment group and sex as factors.

The subject's grade point average, class standing, and study major were included as covariates in the analysis. In the event the ANOVAs revealed significance for either a main effect or the interaction, Scheffe's test of multiple comparisons was used to identify the source of the significance. The acceptable level of significance was set at $\alpha = .05$ throughout the analysis.

Intended Learning

Hypothesis one: There is no significant difference in cognitive learning outcomes of objective relevant (intended) knowledge in an audiovisual presentation as a result of the insertion of learning objectives.

$$H_{o_1} : \sigma E_j = \sigma E_c = 0$$

where E_j = experimental groups 1, 2, and 3 and E_c = experimental control.

The null for research hypothesis one was rejected for both the immediate and delayed post-tests. Intended learning was significantly greater at the .05 level when learning objectives were present. There was virtually no variation in scores attributable to sex differences, and there were no significant interactions between treatment and sex.[31] The covariate of class standing were also not significant. However, the grade point average (GPA) of the subjects was a significant covariate with the intended learning score for both the immediate and the delayed post-test.

Since no sex differences were indicated in the analysis, sex was deleted from all subsequent analysis of the intended learning variable.

Incidental Learning

Hypothesis two: There is no significant difference in cognitive learning outcomes of incidental knowledge in an audiovisual presentation as a result of the insertion of learning objectives.

$$H_{o_2} : \sigma E_j = \sigma E_c = 0$$

where E_j = experimental groups 1, 2, and 3 and E_c = experimental control.

The null for research hypothesis two was accepted for both immediate and delayed tests, but the trend of the results was the reverse of the expected direction. Although not statistically significant ($\alpha = .12$ and $\alpha = .10$ level for immediate and delayed scores respectively), incidental learning tended to increase when learning objectives were present. This trend was present whether the learning objectives were presented before the presentation (advance organizers) or after the presentation (post organizers), but was not present when the learning objectives were interspersed within the presentation (adjunct organizers). On the contrary, incidental learning was slightly lower for the adjunct organizers. As in the intended learning outcomes, sex differences were not significant nor was class standing significantly related to the incidental learning scores. As with intended learning, grade point average was a significant covariate ($\alpha = .001$) with the incidental learning scores and appeared to account for much of the variance for both the immediate and delayed incidental post-test observations.

Location of Objectives

Hypothesis three: Inserting learning objectives at the beginning, at the end, or interspersed during an audiovisual presentation has no significant difference in cognitive learning outcomes of either intended or incidental knowledge.

Since the analysis of variance for incidental learning on both the immediate and delayed post-tests showed no significance for treatment main effects, further testing between groups was not warranted. Consequently, the null hypothesis for the incidental knowledge portion of hypothesis three was accepted by default, i.e., the location of learning objectives has no significant effect on incidental learning performance.

For the intended knowledge portion of hypothesis three, the significance of the differences in performance was tested using the Scheffe method:

$$H_{o_{3a1}} : \mu K_{j_1} = \mu K_c = 0$$

$$H_{o_{3a2}} : \mu K_{j_2} = \mu K_c = 0$$

$$H_{o_{3a3}} : \mu K_{j_3} = \mu K_c = 0$$

where K_j = intended knowledge for experimental groups 1, 2, and 3; and K_c = intended knowledge for experimental control.

The null hypothesis was rejected for the advance organizer experimental group. Placing the learning objectives at the beginning of the audiovisual presentation did significantly improve the intended learning performance on both the immediate and delayed post-test. The null hypotheses for use of the learning objectives as a post organizer and as adjunct organizers were accepted. Although the direction of intended learning performance was greater for both of these treatments than for the control group receiving no learning objectives, the results were not significant at the .05 level.

The Effects of Grade Point Average

Since the grade point average had shown a significant relationship in both the intended and incidental learning outcomes analysis, it was decided to test for possible interaction effects between the treatment and GPA. Specifically, the question arose as to whether the presence and location of learning objectives had a differential effect on high and low achievement students as measured by their overall class standing.

To test for the interaction between treatment and GPA, a two-way analysis of variance was conducted with both the immediate and delayed intended and incidental learning scores. The A and B subjects were grouped as high GPA students, and subjects with C and D grades for the course were designated as low GPA students. It was hypothesized that high GPA students would possess superior organizing skills as evidenced by their demonstrated ability to master the curriculum content of the course. Therefore, the introduction of learning objectives as a device for organizing the intended learning material of the audiovisual presentation would have less beneficial effect. Students lacking highly developed internalized organizing skills, on the other hand, should find the learning objectives more facilitative in learning the objective relevant material presented in the slide-tape program but not in learning the incidental knowledge presented.

While intuitively reasonable, the data did not support this assumption. No significant interactions were present. The grade point average varied as scores varied, but not differentially for treatments.

Learning Retention

In examining the means for the immediate and delayed post-test scores, differences were noted in the amount of learning loss between the treatment groups. This was particularly noticeable for incidental learning results where the rank order of groups three and four were reversed between the immediate and delayed post-test. Ellis points out that the retrieval of information from memory is not independent of how it is stored. He states, ". . . if information is stored in an organized and systematic fashion, it has a better chance to be retrieved."[32]

To test the effects of the use and placement of learning objectives on the rate of learning decay of the intended and incidental knowledge, an analysis of difference scores was made. A paired difference variable was computed for each subject using a repeated measures loss score of $D = X_1 \quad X_2$ for each knowledge type. A two-way analysis of variance was conducted with treatment and sex as the independent variables. The results of the analyses indicated there were no significant differences in learning loss between the treatment groups or the sexes for either intended or incidental knowledge. Although no conclusions can be drawn, it is interesting that the group receiving learning objectives at the beginning of the presentation as advance organizers had the highest retention level for intended knowledge and the lowest retention rate for incidental knowledge, whereas the control group that received no learning objectives and the post organizer group that received the learning objectives at the end of the presentation had the better retention performance for incidental knowledge.

It is emphasized there was no significance in the analysis of differences in the loss of learning between treatment groups; therefore, any conclusions from these data would be unwarranted and misleading.

Affective Variables

Ten items were included in the immediate post-test that were designed to measure by means of semantic differential scales the attitudes of the subjects with regard to the content material and the manner of presentation. The variable names and source of data used in the analysis are presented in Table 2 (p. 15) under "Affective domain" and "Evaluation of presentation." A factor analysis procedure was used to reduce the number of factors that could be used as source variables for testing the variance between treatment groups.

As a result of the factor analysis, the ten variables were restructured into three groups that were defined for further analysis. The variable "optimism" was retained as a unique factor. The variables "importance" and "bias" were combined as a new factor "content evaluation." The remaining seven variables were grouped into a single factor designated "presentation evaluation." The means for these new generated factors by treatment group are displayed graphically in Figure 3.

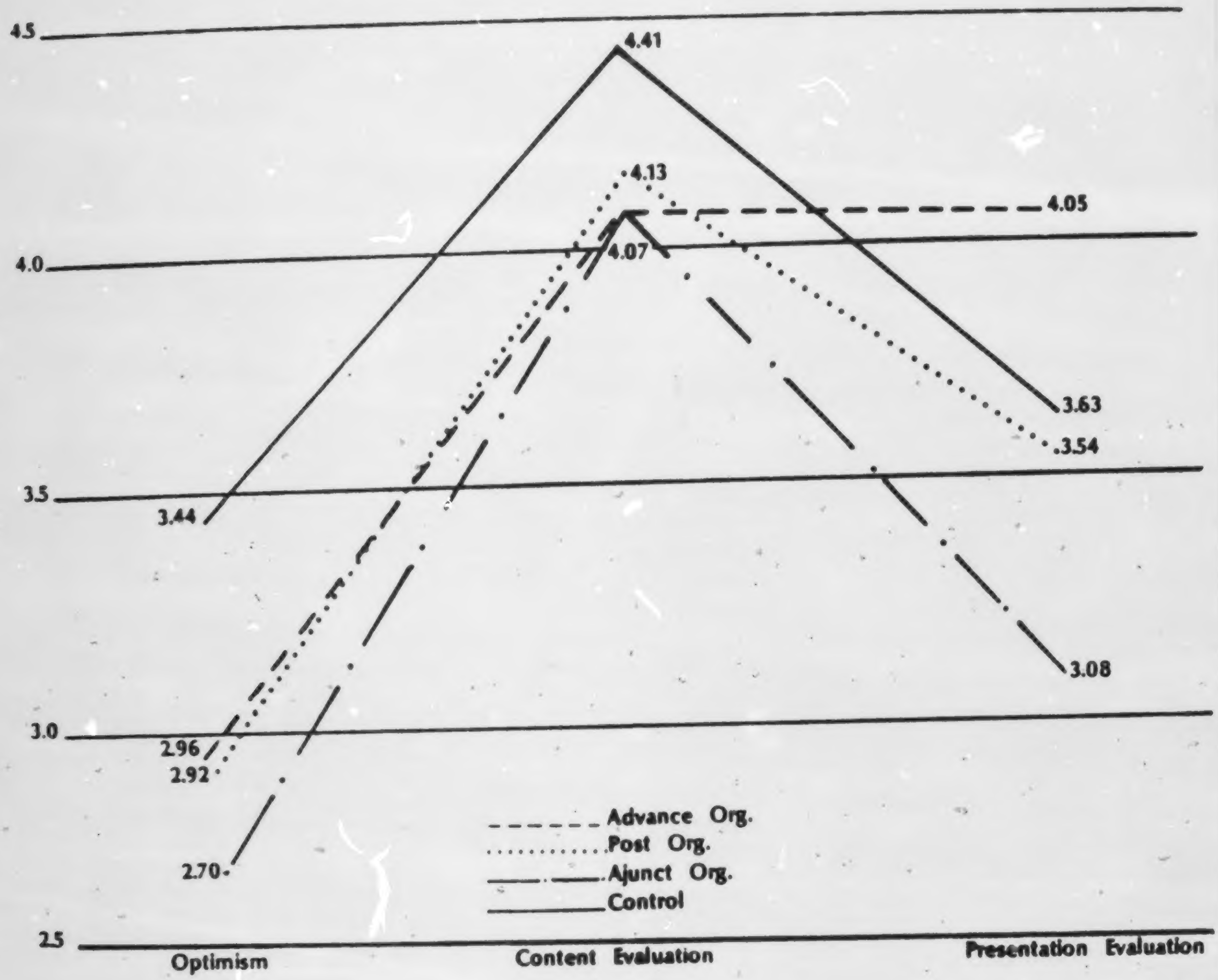


Figure 3

Pattern of Response on Affective and Evaluation Factors

Scores derived from semantic differential scale where 5 = high, 1 = low.

The optimism variable measured the subject's feelings regarding the effects the new communication technology would have on the quality of our future life. The scale ranged from "greatly enhanced" (5) to "greatly degraded" (1). Generally, the treatment groups appeared more pessimistic than the control group with the group receiving the learning objectives as adjunct organizers being the most pessimistic. The differences were not dramatic, however, since less than one point on the five point scale (high = 3.44, low = 2.70) separated the extremes.

The range for treatment group means was even less on the content evaluation factor (high = 4.41, low = 4.07). All of the experimental groups considered the subject matter of importance in that communication media would have an expanding role and an increasingly dominant impact on society. They also perceived the presentation as being relatively unbiased in presenting a balanced treatment of the dangers and benefits of the new media technology as evidenced by the four plus mean ratings on a five point scale.

The greatest range between the experimental groups occurred in the evaluation of the presentation. Each subject rated the presentation on how they liked it overall, how appropriate it was for the course, the content organization, the length, the quality of the visuals presented, whether they would have preferred a different format, and on how interesting it was compared to other presentations. Although the mean rating for all four groups was generally favorable, there was almost a full point difference between the highest mean of 4.05 for the group receiving the learning objectives at the beginning of the presentation and the lowest mean of 3.08 for the group receiving the learning objectives as adjunct organizers throughout the presentation. It appears that inserting the learning objectives during the presentation may have disrupted the content organization and created some annoyance in the viewers.

To test the significance of the differences between the experimental groups' affective and evaluation scores, two-way analyses of variance were conducted. The independent variables were treatment and sex and the dependent variables were the optimism, content evaluation, and presentation evaluation variables generated by the factor analysis.

The analysis indicated there were no significant differences in the subjects' scores on optimism and content evaluation between the experimental groups. Female participants were significantly more pessimistic about the future degradation of the quality of life through the improper exploitation of new communication technology than were the male subjects. However, since there was no significant interaction with the treatment, the result was of little interest to this study.

There was no significant variance in content evaluation scores attributable to either treatment or sex differences.

The evaluation of the presentation itself did, however, show significant treatment effects. Sex had no appreciable effects on the presentation evaluation scores. A Scheffe test of the differences in experimental group means verified the difference between the advance organizer and adjunct organizer group was significant at the .05 level. None of the other comparisons was significant.

The results of all statistical analyses are summarized in Table 4.

Discussion

The findings of this study generally support and extend Ausubel's theory of advance organizers facilitating learning. They are consistent with results obtained by Kaplan,[33] Kaplan and Simmons,[34] and Keyser[35] regarding the efficacy of using learning objectives as advance organizers with prose materials. The length of treatment used in this experiment was relatively short, but since

Table 4

Summary of Statistical Analysis

Independent Variable	Dependent Variable				Analysis Method
	Immed. Int. Learning	Immed. Inc. Learning	Delayed Int. Learning	Delayed Inc. Learning	
Experimental Factors					
Treatment	Sig.[a]	N.S.[b]	Sig.[a]	N.S.[b]	ANOVA
Sex	N.S.	N.S.	N.S.	N.S.	ANOVA
Treatment x Sex	N.S.	N.S.	N.S.	N.S.	ANOVA
GPA	Sig.	Sig.	Sig.	Sig.	Covariate
Treatment x GPA	N.S.	N.S.	N.S.	N.S.	ANOVA
College Major	N.S.	N.S.	N.S.	N.S.	Covariate
Class Standing	N.S.	N.S.	N.S.	N.S.	Covariate
Retention					
	Int. Learning Loss		Inc. Learning Loss		
Treatment	N.S.		N.S.		ANOVA
Sex	N.S.		N.S.		ANOVA

Comparison Tests (Scheffe)[c]

Dependent Variable	
Intended Learning (Immediate and Delayed Post-test)[d]	Adv. Org. > Control Adv. Org. > Adj. Org.
Presentation Evaluation	Adv. Org. > Adj. Org.

---

[a] Test for hypothesis one: null rejected.

[b] Test for hypothesis two: null accepted.

[c] Only those comparisons significant at the .05 level are shown.

[d] Test for hypothesis three: null rejected for intended learning; null accepted for incidental learning.

most audiovisual presentations used as curriculum support are of relatively short duration (one hour or less in length), it would appear that the efficacy of learning objectives used as an advance organizer has strong support.

Locating the learning objectives at the conclusion of a slide-tape presentation or interspersed throughout with the relevant content did not have a significant effect on improving either the intended or incidental learning outcomes. Although the mean scores of the post organizer group were higher than those for the group receiving no learning objectives, the lack of significance limits the support of this study for the use of learning objectives for mathemagenic activities as defined by Rothkopf.[36] Previous attempts to extend the mathemagenic activities hypothesis to audiovisual media have been ambiguous. Dayton concluded that the use of either inserted or grouped post questions would increase intended learning, but that ". . . the use of either inserted or grouped factual knowledge post-questions in fixed-pace, fixed-sequence presentations would not, by itself, ensure an increase in incidental learning."[37]

Based on the results of the present study, it appears the use of learning objectives as post organizers or adjunct aids are what Rothkopf called "mathemagenic neutral," that is, they have no observable effect on the student's ability to retain new information.[38]

A possible explanation of the lack of facilitation of learning when the learning objectives were used as adjunct organizers lies in the nature of the medium. The audiovisual program used for the stimulus response was well organized, tightly structured, with a high degree of aesthetic quality. The insertion of the learning objectives during the presentation could have been very intrusive for the viewer, interrupting the narrative and visual flow of the production and generating frustration or anger. This could have served as an inhibiting condition for the arousal factors and attention devices of the highly professional production values contained in the presentation.

This hypothesis is supported by the results of the presentation evaluation scores. In general, the adjunct organizer treatment group liked the presentation least. Their scores were the lowest on every presentation evaluation item, and they were significantly lower than all three experimental groups for organization. Overall, the advance organizer treatment group liked the presentation more than the other experimental groups. Their evaluation scores were significantly higher than the other three experimental groups for organization, quality of the visuals, and format. In addition, the advance organizer treatment group rated the presentation significantly higher than did the adjunct organizer treatment group for overall evaluation, length, and interest.

The lack of significant difference in the incidental learning outcomes is of some importance. It contradicted findings by Kaplan and Burgin[39] that the use of learning objectives decreased incidental learning with both prose and videotaped instructional materials.

The effectiveness of learning objectives to improve intended learning performance from text materials has been well documented in the literature. The general hypothesis has been that the increased learning stems from the use of the objectives by the learner as directions to learn particular topics within a text (Rothkopf and Kaplan,[40] Kaplan and Rothkopf,[41] Kaplan,[42] and Duchastel and Brown[43]). The concept that they serve as orienting stimuli that permit selective attention to the relevant materials of the text was supported by data showing students took more time when learning objectives were provided (for review of relevant material) and that performance on incidental learning items for students not receiving objectives was much greater. This was explained as being the result of the students' reading non-selectively, i.e., giving equal attention to all material when learning objectives are absent.[44]

The results of this study appear to support a different interpretation of how learning objectives are used by subjects when the instructional material is a fixed-pace, fixed-sequence medium.

Looking at the trends in the data and not merely at the significant levels of effects, it appears the use of learning objectives improved learning performance on the immediate post-test for both the advance and post organizer groups. It can be speculated that in these treatments, the learning objectives serve an organizing function by establishing a categorical structure for examining the information in the audiovisual presentation. In the case of the learning objectives inserted at the beginning of the presentation, the information is considered and coded or filed by the viewer under the appropriate category as it unfolds. When the learning objectives are provided at the conclusion of the presentation, the viewer must conduct a mental review of the presentation reorganizing the material according to the new structure. In either case, the process of considering the information and determining objective relevance (intended knowledge) or non-relevance (incidental knowledge) improves recall for the post-test.

The significant relationship between the grade point average of a student and the score obtained on both intended and incidental knowledge was not surprising. Better students should be expected to perform better than poorer students. It was somewhat surprising that no interaction was found, however, since ability to organize material should be one of the factors in learning performance.[45] It would seem reasonable, therefore, that if learning objectives provide an organizing function they would be more beneficial to those students without highly developed internalized organizing skills, i.e., the poorer students. The lack of interaction indicates this did not happen.

Conclusions and Implications

The most substantive finding was that presentation of learning objectives as an advance organizer facilitated the intended learning for a slide-tape audiovisual presentation. Of particular interest was that these findings were also observed after a two-week delay. Thus, the facilitative effects of prior presentation of learning objectives was found for delayed as well as immediate recall.

Although no significant effects were found for the facilitation of incidental learning, the acceptance of the null hypothesis was of considerable importance for instructional strategy. The conclusion drawn is that instructors may present learning objectives prior to audiovisual presentations with the expectation they will enhance the learning of the objective relevant material without inhibiting the amount of non-objective relevant material learned. From the direction of the data, it appears the tendency was for the learning objectives to improve rather than lessen incidental learning outcomes.

Also of importance for instructors and producers were the findings that the use of the learning objectives as an advance organizer had no significance for the affective variables measured and they did not detract from the students' evaluation of this presentation. To the contrary, the trend of the data indicated the students receiving the learning objectives prior to the presentation liked the presentation more than when no learning objectives were used.

It would be unreasonable to expect that the producers of commercial audiovisual educational materials would provide learning objectives as an integral part of their presentations. Such materials are designed for the widest possible market and the specification of objectives might limit the application of the material or, if not, be so vague and encompassing as to negate their effectiveness. In addition, there is a great deal of audiovisual material (including television programing) that can be used for curriculum support although it is not designed for educational purposes.

Learning objectives can be developed by instructors for the audiovisual materials they use. The insertion of the learning objectives prior to the presentation of the audiovisual program can be accomplished rather easily. If the primary concern of the user of audiovisual materials is to increase intended learning, the insertion of learning objectives prior to the presentation is recommended.

Reference Notes

[1] K. P. Komoski, "EPIE Institute's Approach to the Nonprint Materials Information Problem," *Nonprint Media Information Networking: Status and Potential*, ed. James W. Brown. U.S., Educational Resources Information Center, ERIC Document NIE-C-74-0027, August, 1976, pp. 40-43.

[2] There are literally hundreds of media catalogs available to curriculum developers, planners, and educators. Most of these catalogs are based on institutional library holdings, commercial producer and distributors' inventories, or special areas of professional interest. An important effort at comprehensive inventories are the *National Information Center for Educational Media (NICEM) Indexes* (Los Angeles: The University of Southern California, 1976). These data bases contain nearly 500,000 title listings.

[3] D. P. Ausubel, *Educational Psychology: A Cognitive View* (New York: Holt, Rinehart and Winston, 1968).

[4] R. M. Gagne, *The Conditions of Learning* (New York: Holt, Rinehart and Winston, 1965).

[5] E. Z. Rothkopf, "Learning from Written Instructive Materials: An Exploration of the Control of Inspection Behavior by Test-like Events," *American Educational Research Journal* 3:241-249 (1966).

[6] B. F. Skinner, *The Behavior of Organisms: An Experimental Analysis* (New York: Appleton-Century-Crofts, 1938).

[7] D. P. Ausubel and F. G. Robinson, *School Learning: An Introduction to Educational Psychology* (New York: Holt, Rinehart and Winston, 1969), p. 315.

[8] D. P. Ausubel, *The Psychology of Meaningful Verbal Learning* (New York: Grune and Stratton, 1963).

[9] D. P. Ausubel, "A Cognitive Structure View of Word and Concept Meaning," *Readings in the Psychology of Cognition*, ed. R. C. Anderson and D. P. Ausubel (New York: Holt, Rinehart and Winston, 1965), pp. 58-75.

[10] Ausubel and Robinson, p. 315.

[11] *Ibid.*, p. 316.

[12] *Ibid.*

[13] R. M. Gagne, "Educational Objectives and Human Performance," *Learning and the Educational Process*, ed. J. D. Krumboltz (Chicago: Rand, McNally, 1965), p. 9.

[14] K. C. Garrison and R. A. Magoon, *Educational Psychology: An Integration of Psychology and Educational Practices* (Columbus, Ohio: Charles E. Merrill, 1972), p. 298.

[15] Garrison and Magoon, p. 309.

[16] H. C. Ellis, Fundamentals of Human Learning and Cognition (Dubuque, Iowa: William C. Brown, 1972), pp. 130, 132.

[17] E. Z. Rothkopf, "Learning from Written Instructive Materials," p. 247.

[18] Ibid., p. 248.

[19] E. Z. Rothkopf, "Experiments on Mathemagenic Behavior and the Technology of Written Instruction," Verbal Learning Research and the Technology of Written Instruction, ed. E. Z. Rothkopf and P. Johnson (New York: Teachers College Press, 1971), pp. 284-303.

[20] For a partial review, see P. C. Duchastel and P. F. Merrill, "The Effects of Behavioral Objectives on Learning: A Review of Empirical Studies," Review of Educational Research 43:53-69 (1973).

[21] R. Kaplan, "Effects of Grouping and Response Characteristics of Instructional Objectives When Learning from Prose," Journal of Educational Psychology 68:424-430 (1976); R. Kaplan and E. Z. Rothkopf, "Instructional Objectives As Directions to Learners: Effects of Passage Length and Amount of Objective-Relevant Content," Journal of Educational Psychology 66:448-456 (1974); E. Z. Rothkopf and R. Kaplan, "An Exploration of the Effect of Density and Specificity of Instructional Objectives on Learning from Text," Journal of Educational Psychology 63:295-302 (1972).

[22] P. C. Duchastel and B. R. Brown, "Incidental and Relevant Learning with Instructional Objectives," Journal of Educational Psychology 66:448-456 (1974).

[23] R. Kaplan and F. G. Simmons, "Effects of Instructional Objectives Used as Orienting Stimuli or As Summary/Review upon Prose Learning," Journal of Educational Psychology 66:614-622 (1974).

[24] R. H. Bruning, "Effects of Review and Testlike Events within the Learning of Prose Materials," Journal of Educational Psychology 59:16-19 (1968).

[25] L. T. Frase, "Learning from Prose Material: Length of Passage, Knowledge of Results and Position of Questions," Journal of Educational Psychology 58:266-272 (1967).

[26] Rothkopf, "Learning from Written Instructive Materials."

[27] A. Lumsdaine, M. A. May, and R. S. Hadsell, Learning from Films (New Haven: Yale University Press, 1958), pp. 84-106.

[28] D. K. Dayton, "The Effects of Inserted Post-Questions upon Learning from Slide-Tape Presentations: A Test of the Mathemagenic Hypothesis" (unpublished Doctoral dissertation, Indiana University, 1976). Ann Arbor: University Microfilms, 1977, No. 77-1881.

[29] Media and Meaning: Human Expression and Technology (White Plains, New York, The Center for Humanities, Inc., 1973).

[30] F. G. Brown, Principles of Educational and Psychological Testing (2nd ed.; New York: Holt, Rinehart and Winston, 1976), p. 96.

[31] The SPSS program used for analysis of variance eliminates two-way and higher interactions when the interaction sum of squares approaches zero. In this situation, the inversion of the cross-product matrix "fails" because of the minimum sum of squares value. In the procedure used for this analysis, the factor and covariate effects are assessed concurrently, then interactions are examined (N. H. Nie and others, SPSS (Statistical Package for the Social Sciences) (2nd ed.; New York: McGraw-Hill, 1970), pp. 416-417). The assessment of effects is:

--Covariates: adjust for factors and all other covariates;
--Factors: adjust for covariates and all other factors;
--Interactions: adjust for covariates, factors, and all other interactions of the same or lower orders.

[32] H. C. Ellis, Fundamentals of Human Learning and Cognition, p. 121.

[33] R. Kaplan, "Effects of Grouping."

[34] R. Kaplan and F. G. Simmons, "Effects of Instructional Objectives."

[35] D. F. Keyser, "Relationship of Placement of Performance Objectives and Reading Comprehension Levels on the Learning of Intentional and Unintentional Material in Written Prose" (unpublished Doctoral dissertation, University of Michigan, 1976). Ann Arbor: University Microfilms, 1976, No. 76-19, 165.

[36] E. Z. Rothkopf, "The Concept of Mathemagenic Activities," Review of Educational Research 40:325-336 (1970).

[37] D. K. Dayton, "The Effects of Inserted Post Questions."

[38] E. Z. Rothkopf, "Concerning Parallels Between Adaptive Processes in Thinking and Self-Instruction," Approaches to Thoughts, ed. J. F. Voss (Columbus: Merril, 1969), pp. 299-316.

[39] R. Kaplan and E. M. Burgin, "Instructional Objectives As an Aid to Learning from Prose vs. Videotaped Instruction" (unpublished paper, R. Kaplan, P.O. Box 12, Mountain View, California, 1976). (Mimeographed)

[40] E. Z. Rothkopf and R. Kaplan, "Exploration of the Effect of Density."

[41] R. Kaplan and E. Z. Rothkopf, "Instructional Objectives."

[42] Kaplan, "Effects of Grouping."

[43] P. C. Duchastel and B. R. Brown, "Incidental and Relevant Learning."

[44] Kaplan and Burgin.

[45] H. C. Ellis, Fundamentals of Human Learning and Cognition, p. 132.

A Comparison:
Predicting College Level Academic Success With Tests of Cognitive Style and Cognitive Aptitude

Dr. Thomas M. Schwen, Director

Anne K. Bednar, Project Director
Division of Development and Special Projects
Audio Visual Center, Indiana University
Bloomington, Indiana

A Comparison:
Predicting College Level Academic Success
With Tests of Cognitive Style and Cognitive Aptitude

This study attempted to establish the relationship of cognitive aptitudes and styles to measures of general and more subtle academic achievement. It was expected that the associations of aptitudes and styles to two general measures of achievement would converge to form similar correlational patterns; associations of aptitudes to more subtle measures of achievement would also converge, while associations of styles to subtle achievement variables would diverge. Relationships with the subtle achievement measures were as expected, but the associations with general achievement measures did not converge. It is suggested that future style research focus on simpler variables theoretically linked to mediational process in the style literature.

Introduction

In the systematic development of instruction, one major concern of research into individual differences has been the attempt to relate psychological theory to the prediction of college level academic success and ultimately to the effectiveness of classroom instruction. Initially such prediction was used for the purpose of selection; more recently, with the development of aptitude treatment interaction (ATI) research, aptitude measurement has been used to optimize learning for all students through differential application of instructional treatments to suit various individual differences.

The process of instructional design and development is especially concerned with this relationship of aptitudes and treatments when it deals with the development of academic courses in higher education. A variety of course designs such as lecture, recitation, Keller Plan, contract or tutorial instruction may be employed separately or in combination as the organizational framework of course format or presentation. A range of learner aptitudes such as quantitative, verbal, spatial or problem solving abilities may interact with content or course design options to facilitate the success of certain learners. For example, it has been shown that the personality variables of achievement via conformance and achievement via independence interact with certain course design factors so that students who score high on a measure of conforming motivation demonstrate higher achievement in settings which reward conforming behavior while students who score high on a measure of independent motivation demonstrate higher achievement in course designs which reward independent behavior (Domino, 1968).

Problem

Because of the possibility of such interactions between learner characteristics and elements of course design, it is important to the field of instructional development to investigate whether achievement in a specific instructional environment (especially one with a strong technological design) correlates with academic success or whether certain learner variables are associated with achievement in specific instructional designs.

Traditionally research has examined the effects of cognitive aptitudes and personality dimensions on learning. Recently interest has grown in a new category of variables, cognitive style, described as an individual's characteristic or preferred way of perceiving and/or organizing the world (Messick, 1976; Ausburn & Ausburn, 1978). In relation to models of cognitive aptitudes and personality dimensions, cognitive styles are represented theoretically as mediating processes which cut across the somewhat arbitrary aptitude and personality categories (Messick, 1976; Royce, 1973). A number of authors have observed that this class of variables, cognitive styles, should be useful in predicting academic success (Messick, 1976; Cross, 1976; Ausburn & Ausburn, 1978). In addition, because cognitive styles characterize the way in which an individual learns rather than the what of learning, they should constitute an especially valuable variable in the study of learning in relation to variations in course designs. With theoretical relevance to both the processes and the outcomes of academic performance, cognitive style may contribute to the prediction of success and/or differential behavior in a variety of learning environments.

The basic association between cognitive style and the traditional research variables of cognitive aptitudes and personality dimensions, especially as they relate to the prediction of academic success, however, is largely unexplored and, therefore, an important point of departure for research. This study approached the fundamental question: In the prediction of academic achievement, both general (Grade Point Average) and at the level of course design, do a selected class of cognitive style measures add to the variance accounted for by traditional cognitive aptitude measures? This general research question was reduced to the following more specific questions:

1. Do cognitive style measures contribute to the prediction of general academic achievement (GPA)?
2. When considered together do cognitive style measures predict general academic achievement over and above traditional cognitive aptitude measures?
3. Do cognitive style measures contribute to the prediction of academic achievement in a specific course design?
4. When considered together do cognitive style measures predict academic achievement in a specific course design over and above traditional cognitive aptitude measures?
5. What is the shared variance between style and cognitive aptitude measures in predicting academic achievement measures?
6. Do patterns of prediction variables for general academic success differ from patterns of prediction variables for academic achievement in a specific course design?

Related Research

Other than with Witkin's (1977) construct Field Dependence-Field Independence, there has been a reasonably modest amount of research examining the validity of cognitive style by academic relationships. Witkin's review of research is quite extensive; the major generalizations relevant to this study are:

1. Field Dependent students seem to attend to and learn more effectively from social cues.
2. Field Independent students are less sensitive to social punishment and they seem to learn more in the absence of external reward, under intrinsic motivating contingencies.
3. Field Independent students seem to use more mediational processes such as the analysis and restructuring of stimuli, especially in unstructured instructional environments.

The following style dimensions were included as variables in the study (see Messick (1977) or Ausburn & Ausburn (1978) for a more extensive review):

1. Field dependence-field independence is a measure of an individual's self, non-self segregation as reflected by reliance on internal versus external cues in relating self to the surrounding environment. Behaviorally, field dependence influences the extent to which an individual operates autonomously in interpersonal relations and the manner in which one processes information from the environment. The test used to measure field dependence-field independence was Witkin's Group Embedded Figures Test. It involves disembedding a simple figure from a complex pattern.

2. Closure flexibility measures an individual's ability to keep a configuration in mind in spite of distraction. There is some claim that it is a measure of temperament and that it differentiates industrial occupation groups. The test which measures closure flexibility, the Concealed Figures Test, is very similar to the Group Embedded Figures Test and has been used as a measure of field dependence-field independence.

3. Visualization and spatial relations were measured using Flags and the space relations scale of the Differential Aptitude Test. Flags involves the rotation and matching of a two dimensional design while the DAT requires mental folding of a two dimensional pattern to form a three dimensional shape.

4. Conceptual differentiation is an individual's characteristic tendency to employ relatively more or fewer distinctions between concepts in category formation on a free categorizing task. The Clayton-Jackson Object Sort, used to measure conceptual differentiation and compartmentalization, is a free categorizing task in which individuals group fifty objects into an unspecified number of categories and label the logical reason for category membership.

5. Compartmentalization reflects an individual's tendency to conceptualize objects in discrete, rigid categories. It is assumed to reflect an inability to produce diverse ideas. The measurement of compartmentalization is a count of the number of objects which an individual is unable to incorporate into categories in the Clayton-Jackson Object Sort.

6. Conceptualizing Style indicates an individual's characteristic manner of establishing criteria for category formation, either on the basis of physical, inferred or relational similarity (such as two dissimilar objects each related to a third). The measure of conceptualizing style was a panel judgment of the category reasons on the Clayton-Jackson Object Sort.

7. Leveling-sharpening refers to an individual's mode of organizing a sequence of stimuli as either relatively more or less highly articulated. The Schematizing Test consists of a series of ninety squares varying in size from one to eighteen inches. As the squares are projected on a blank screen in a darkened room, the individual is asked to judge the size of the squares in inches. A "sharpener" tends to make more distinctions between stimuli of similar but unequal size while a "leveler" tends to group similar sizes into a single category reporting unequal squares as the same.

8. Category width or equivalence range describes whether an individual tends to form broad, inclusive categories or narrow, exclusive categories. The instrument used to measure category width is the Pettigrew Category Based Width Scale. The individual is asked to estimate the largest and smallest example of an item in existence when given an indication of an average size. Broad categorizers will choose extremes while narrow categorizers will estimate closer to the average.

9. Complexity-simplicity measures the capacity of an individual

to view the environment, especially the social environment in a multidimensional way. Kelly's Role Construct Repertory Test, modified for analysis via multidimensional scaling served as the measure of complexity-simplicity. Participants were asked to rate fifteen paired roles as similar or dissimilar to each other. Complex individuals made finer distinctions between roles, using more dimensions to differentiate.

10. Constricted versus flexible control describes the extent to which an individual is able to restrict his/her attention and control susceptibility to distraction. The test which measures this style is the Speed of Color Discrimination Test. In the first part of the test participants are to identify the color of ink used to print short series of dots. The test is highly speeded; participants are given only fifty seconds to identify the color of ninety-six patterns. On the second part, instead of dot patterns, the name of a color is written in a different color ink. The individual is to identify the color of the ink. More flexible individuals are less distracted so that the difference in their scores on the two parts will be less.

Method

The college level course from which the participants for this study were drawn was a large-enrollment introductory Geography course for freshmen and sophomores at a large mid-western university. The course design was an auto-tutorial mastery design which has been rigorously developed over a period of five years.

Instructional environments consisted of a laboratory setting in which students utilized behavioral objectives, audio tapes, various visual stimulus materials and practice tests in a typical mastery pattern. A discussion section was used for motivational exercises (e.g. educational games) and clarification of content. Previous evaluation efforts have shown (Schwen and Keller, 1977) that the treatment is quite demanding, eliciting many productive learning behaviors.

Course grades were based on three objective exams and a judged term paper. The exams were rigorously developed in the process of course development. Most items (twenty out of thirty) were of a concept learning form while the paper was a modified problem solving task.

Participants in the study consisted of 240 undergraduate students, two-thirds of whom were enrolled in the course during the process of the study and one-third of whom were a random sample of students who had taken the course in prior semesters. These groups of students were compared; there were no major differences between the two groups. This group of students has been shown in previous research to be similar to the general population of college freshmen and sophomores (Schwen and Keller, 1977).

These students were tested using conventional large group testing procedures. The group of students who had taken the course in past semesters were tested in two, three-hour evening sessions while students enrolled in the course were tested in a series of nine, forty-minute testing sessions spread over a fifteen week span. The order of tests was varied to reduce the effects of fatigue and practice as well as any interactive effects between measures. Those subjects for whom test data was incomplete were dropped from the study.

Tests included measures of cognitive aptitude, cognitive style and academic success. SAT scores and a ratio variable of high school rank to class size were used as traditional or "markers" in cognitive aptitude predictors of academic achievement.

Cognitive style constructs were selected on the basis of factor analytic studies (Royce, 1973; Gardner, et. al., 1968). The criteria used to select the measures were that measures should be reasonably well defined and coordinate; independent of one another; capable of being administered in large groups; reasonably short; appropriate for college-aged participants and reliable.

Measures of the dependent variable, academic success, were of both a general and a course specific nature. The general measure of academic success was the students' first semester grade point average. Course specific measures of academic success included the second objective test and a term paper. The tests and processes of judging papers were developed along with the course over a five year period and reliabilities exceed traditional standards for both measures (.85 - thirty item test; .92 - inter-rater reliability on paper).

Design

This study was the first of several establishing the relationships between style and instructional outcomes. The dependent variables were selected to represent the full range of performance from complex constructs such as Grade Point Average (GPA) to concept and problem solving behaviors.

Traditional cognitive predictors of achievement were used as "markers" or bench marks for the observed relationships. This use of traditional predictors made it possible to compare the utility and explanatory power of adding style constructs to predictors of achievement.

Some criticisms of style research have suggested these measures are confounded with general cognitive ability measures. The use of multiple stepwise regression procedures permitted a partial test of this criticism. The overlap between style and cognitive measures was examined with each prediction of a dependent measure.

Our expectations regarding the data were that the more complex composite variables, GPA and the total points in the course, would illustrate similar patterns of relationships in combination with the aptitude measures. This convergent expectation or hypothesis was based on the assumption that these variables were roughly equal in complexity and a result of the same complex processes. Also, we did not expect style measures to be highly correlated with these dependent variables. We reasoned that these variables were quite different in complexity and the theory of the style measures (Messick,1977) would suggest that the style measures are bi polar mediational processes where the same outcomes may be achieved through the different (bi polar) processes.

Our expectations regarding the more subtle course dependent variables and the cognitive styles and aptitudes were that the cognitive aptitudes were so complex that they would converge in association with the more subtle dependent variables. In other words, general academic aptitude would not differentiate between more subtle measures of academic achievement. However, we expected the style measures to diverge, or illustrate different correlational patterns, in combination with the dependent variables. Our expectation was that the more subtle style constructs and related measures could differentiate between the subtle measures of academic success. In other words, we assumed the different dependent variables involved different mediational responses.

Data Analysis

The several multiple regression analyses are summarized in Figure 1. The amount of variance shared by the cognitive aptitudes and dependent measures is shown by the solid line joining the ✸ shape. The amount of variance shared by the style measures and dependent measures is shown by the dotted line joining the ○ shape. The amount of variance the style measures shared with the dependent measure after the common variance between aptitude measures and dependent measures was removed from the regression equation is shown by the dash-dot line joining the Δ shape.

The total variance shared among cognitive aptitude and style measures with the achievement measures is recorded at the bottom of the chart.

The calculation of common variance followed the SPSS format (Multiple Regression Subprogram) with adjustment for error.

The individual cognitive style measures that accounted for the greater part of the variance in each equation are indicated with a letter code immediately adjacent to each percent-of-variance data point. The problem solving variable, shown to the right on the graph, is a measure from another study we conducted. These are problems from the classical literature on problem solving. These data were incorporated to demonstrate the contrast in relationships between the more complex multivariate course related variables and the more precise univariate problem solving variable.

The most striking of the several relationships would seem to be the interaction between the complexity of the dependent variable and the amount of unique variance between style and the dependent variables. In general it would seem that the more univariant variables would be most suitable for future research with the style variables.

It should be noted that the dependent variables are not independent. Total is a composite score that includes T_2 (the second objective test) and the problem solving paper. The total score was utilized to compare it with GPA. As noted, the expectation was that these variables would exhibit similar relationships. This is, obviously, not the case. We have no explanation for this discrepancy. The measure is either invalid or there is a complex interaction between the rest of the variables in the composite T_2 and paper and the predictor variables.

The magnitude of the relationships between style and the dependent variables is modest for the course variables. Also, the overlap between style and cognitive variables could be interpreted as support for the previously noted criticism of the style measures. Our interpretation is that the theoretical definitions of these style constructs include mediational or process responses. Many of the definitions are bi polar, not necessarily implying more or less skill but different mediational responses. Therefore, the style by achievement validity coefficients may be expected to be modest. Dependent variables closer to the mediational processes or dependent variables that are instrumental in producing the dependent variables may produce associations of greater magnitude.

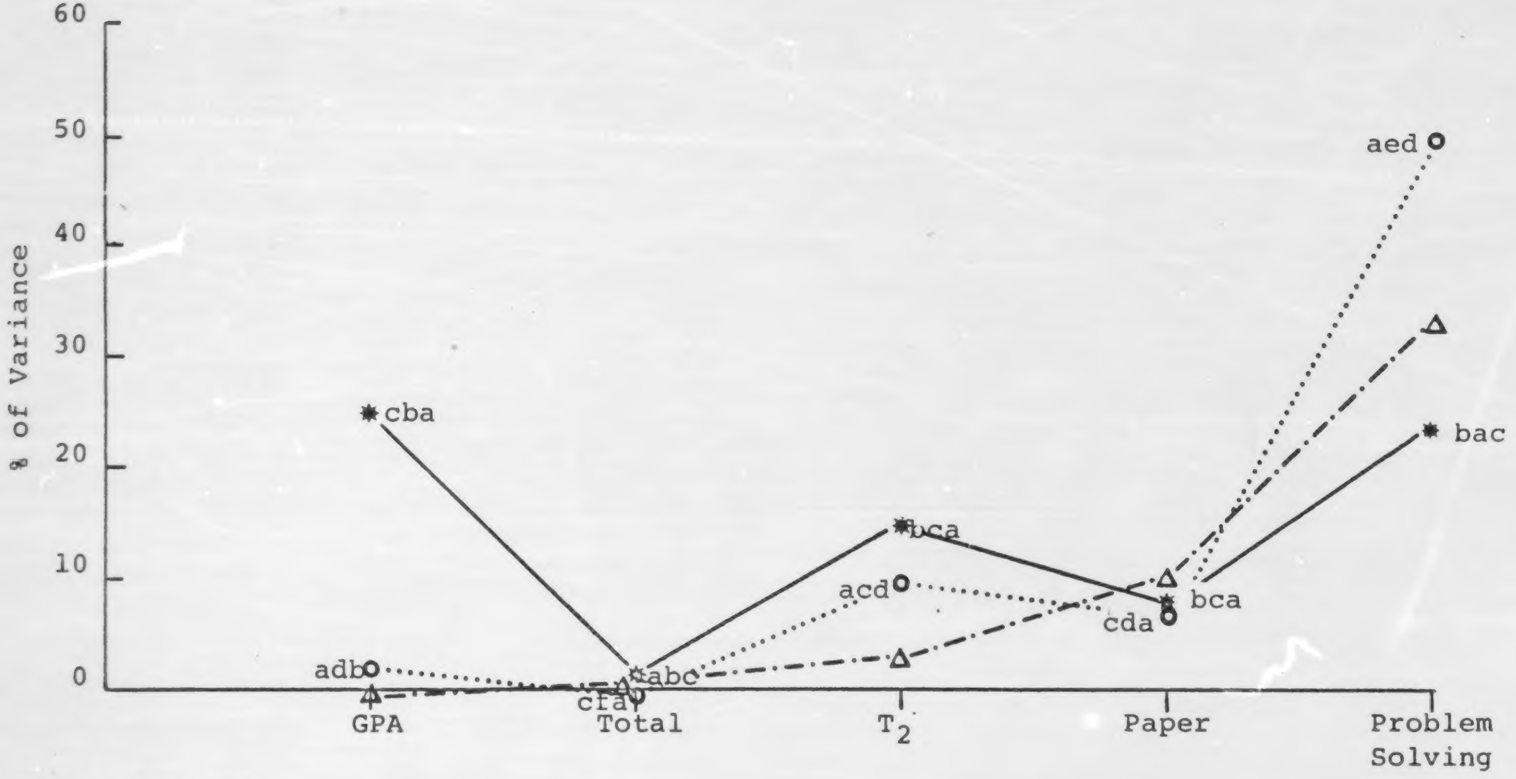
Consistent with this line of reasoning, the rather substantial associations between styles and classical problem solving items may be seen as evidence of the confounding of cognitive spatial aptitude (DAT, Concealed Figures, etc. are sometimes described as spatial aptitudes) and style. This topic is being explored in more depth in a related study.

The order (in the regression equations) of the style and aptitude variables across the dependent variables is of some interest. The concept learning test two (T_2) and the paper seem to be drawing on the same cognitive aptitude but different style variables. This pattern is satisfying because the aptitude constructs are defined as quite gross or complex whereas the style constructs are defined as being more subtle. Also, the dependent variables should be drawing on different cognitive processes. The assumption was that the complex measures would converge and the subtle measures would diverge in association with reasonably subtle differences in dependent variables.

In summary, this study attempted to probe the validity of cognitive style by typical dependent variables in the context of a mastery oriented course. Our expectations were that the associations between complex variables GPA and total points in the course with cognitive aptitudes and styles would converge or exhibit similar correlational patterns. This was not the case, leading us to question the validity of the total measure. Our expectations regarding the more univariant dependent variables, concept learning and paper performance was that the association would converge in respect to aptitude and the dependent measures. In other words there would be similar patterns of correlation across these sets of variables. Further, our expectation was that the style by dependent measure associations would diverge. In other words there would be different patterns of correlation. These expectations were more or less confirmed leading us to conclude the less complex style measures may be more closely linked to the concept and problem solving variables as the theory in these two areas would suggest.

The magnitude and patterns of the relationships between style and the course dependent variables led us to conclude there would be merit in continuing this kind of exploratory effort. However, our new efforts will be focused on simpler theoretical variables that are theoretically linked to the mediational processes described in the theory of cognitive style. In conjunction with our associates we are currently exploring the association between style and instrumental behaviors such as perceptual attack skills, selected discussion and problem solving behaviors.

Figure 1.
Summary Table of Multiple Regression Analyses



Total Variance Accounted for by Cognitive Aptitude and Cognitive Style: .25115 .02296 .19410 .19431 .57443

✱ = a.SAT Verbal, b.SAT Math, c.Class Rank - APTITUDES

o = a.Closure flexibility, b.Compartmentalization, c.Constricted/Flexible, d.Category Width, e.Space Relations, f.Visualization _ COGNITIVE STYLES

△ = Variance accounted for by Cognitive Style after Cognitive Aptitude

References

Ausburn, L.J., & Ausburn, F.B. Cognitive styles: Some information and implications for instructional design. Educational Communication and Technology, 1978, 26, 337-354.

Cronbach, L.J., & Snow, R.E. Aptitudes and instructional methods: A handbook for research on interactions. Irvington Publishers, Inc.: New York, 1977.

Cross, K.P. Accent on learning. Jossey-Bass Publishers, 1976.

Domino, G. Differential predictions of academic achievement in conforming and independent settings. Journal of Educational Psychology, 1968, 59, 256-260.

Gardner, R.W., Lohrenz, L.S., & Schoen, R.A. Cognitive control of differentiation in the perception of persons and objects. Perceptual and Motor Skills, 1968, 26, 311-330.

Messick, S. & Associates. Individuality in learning. Jossey-Bass Publishers, 1976.

Royce, J.R. "The conceptual framework for a multi-factor theory of individuality." in Royce, J.R. (Ed.), Multivariate analysis and psychological theory. New York: Academic Press, 1973.

Schwen, T.M., & Keller, J.M. A case study: Developing convergent formative evaluation methodology. Journal of Instructional Development, 1977, 1(1), 31-35.

Witkin, H.A., Moore, C.A., Goodenough, D.R., Cox, P.W. Field-dependent and field-independent cognitive styles and their eucational implications. Review of Educational Research, 1977, 47(1), 1-64.

ENVIRONMENTAL & PERSONAL FACTORS AFFECTING INSTRUCTIONAL DEVELOPMENT BY THE MEDIA PROFESSIONAL AT THE K-12 LEVEL

by

Philip M. Turner, Ed.D.
Nina N. Martin, Ed.D.

Graduate School of Library Service
University of Alabama
P. O. Box 6242
University, Alabama 35486

January 30, 1979

ABSTRACT

This study investigated the relationship of environmental and personal factors of K-12 media professionals in Alabama and the reported performance by this population of 28 instructional development activities. Conclusions revealed a positive relationship of competencies, attitude, budget, empathy, and non-supervisory time with reported performance of instructional development activities. Also concluded was a negative relationship of parent and student volunteers with the dependent variable. Recommendations include competency testing of media professional preparatory program graduates, research into empathy training, and an increase in non-supervisory time for media professionals.

INTRODUCTION

Instructional development is a systematic approach to the design, production, and utilization of complete systems of instruction (Association for Educational Communications and Technology, 1977: 172). Within this realm of instructional development, six sub-functions are delineated. These include Research-Theory, Design, Production, Evaluation-Selection, Logistics, and Utilization-Dissemination (AECT, 1977: 164).

The role of the K-12 level media professional as instructional developer has been explicated in textbooks for potential media professionals (Brown, Norberg, and Srygley, 1972: 1; Erickson, 1968: 10; Prostano and Prostano, 1971: 214-44; Davies, 1969: 41-58). A Delphi Study by Jetter (1972) predicted the assumption of the role of instructional developers as the major trend in the future for media professionals. Finally, the role of the K-12 media professional as instructional developer was formalized in Media Programs: District and School (American Library Association/AECT, 1975).

Studies have been undertaken attempting to determine the relationship of several variables analogous to instructional development and a wide range of independent variables. These include Schulzetenberge (1970): curriculum development activities with undergraduate major, similarity of interests between media specialists and teacher, and extroversion; Maudus (1975): curriculum involvement with teaching structures; Larsen (1971): role assignments with time, budget, and principal's perception; Leeper (1975): quality of media program with both the attitude of principal and media professional; Daniel (1974): allocation of responsibility to media professional with submissiveness and accomodation on the part of the media professional; and Kerr (1977): acceptance of the role of instructional developer with role-taking ability.

The role of the media professional as instructional developer is a relatively recent phenomenon. While the role of instructional developer is promulgated by national organizations, the recency of a standard definition and description of activities involved in instructional development has precluded in-depth studies.

THE PROBLEM

The literature of library education and educational technology is replete with exhortations for the media professional to become involved in instructional development at the school level. A survey of the literature yields a series of personal and environmental factors that would logically be linked to instructional development. No study was located, however, which attempted to establish a relationship between personal and environmental data and the performance by the media professional at the K-12 level of the instructional development activities as delineated by the AECT definition.

Purposes of the Study

The purposes of the study were:

1. to determine the relationship between personal and environmental variables and the amount of reported instructional development.

2. to determine the effect of the variables on the reported performance of instructional development activities within each of the domains delineated by the AECT definition.

3. to identify the factors which may lead to increased amounts of instructional development by K-12 media professionals.

Limitations

This study was restricted in the following areas:

1. the population was restricted to media professionals employed at the K-12 level in the State of Alabama.

2. the personal and environmental factors were restricted to those covered on the instruments.

METHODOLOGY

This research utilized questionnaires as the data collection instruments. This method was chosen in order to collect a wide range of data on a large population.

Independent Variables

There were 38 independent variables drawn from the environmental and personal factors and combinations of these factors.

Dependent Variable

The dependent variable was the amount of reported instructional development undertaken by the media professional.

Subjects

The subjects were a sample of 300 drawn from K-12 media professionals in the State of Alabama as identified by the State Department of Education. The subjects were chosen through a random number process and represented a cross-section of grades served.

Instrumentation

Five instruments were used including:

1. Media Professional Background Data Questionnaire. This is an investigator-designed instrument which gathered personal information from the media professional. Variables measured by this instrument included:

a. age

b. sex

c. number of years of classroom experience

d. number of years of school library experience

e. number of years at present position

f. possession of bachelors, masters, and/or sixth-year degree

g. years since last coursework

h. number of design courses taken

i. number of traditional library courses taken

j. number of education courses taken

k. number of production courses taken

l. total number of above courses

m. self-rated competency in design of instruction

n. self-rated competency in traditional library skills

o. self-rated competency in education area

p. self-rated competency in production

q. total self-rated competency

2. Environmental Data Questionnaire. This investigator-designed instrument gathered information relating to factors in the working environment which could affect instructional development activity. Variables measured by this instrument included:

a. grade level served

b. number of hours assigned to library

c. amount of time each day not involved in the supervision of students

d. ratio of library media professionals to students

e. location of AV services

f. presence of a production area

g. print budget

h. production supply budget

i. commercial non-print budget

j. total budget

k. number of services provided by district central office

l. number of production equipment items available

m. number of professional journals read

n. amount of teacher planning time/day

o. teacher released time during year for course planning

p. number of paid aides

q. number of student assistants

r. number of parent volunteers

s. number of innovative teaching structures (team teaching, modular scheduling, open-space classrooms, etc.)

3. Empathy Scale. The respondent's empathy state was quantified using the Empathy Scale developed by Robert Hogan (1969). The long form of this instrument consists of 31 items from the California Psychological Inventory, 25 from the Minnesota Multiphasic Personality Inventory, and eight items from experimental testing forms. A test-retest reliability coefficient of .84 has been found, and internal consistency estimates of

.71 have been reported (Grief and Hogan, 1973). The conceptual validity of the scale has been demonstrated in several studies (Hogan and Mankin, 1970; Hogan, 1973, Kurtines and Hogan, 1972; Hogan and Dickstein, 1972).

The short form of the Empathy Scale, which consists of only those 31 items from the California Psychological Inventory, has routinely correlated above .90 with the 64 item version (Grief and Hogan, 1973: 281). The short form was used for this study. The author felt that the use of the simpler form would significantly increase return rate.

4. Media Professional Activities Checklist. This instrument was developed by the investigator and contained 28 activities drawn from the Task Analysis Survey Instrument (ALA, 1969), from other literature, and from correspondence with professionals in the field. The respondent checked a Likert-type scale ranging from 0 to 4 for each of the 28 activities. Anchors for these were "never" and "frequently." Total possible scores ranged from 0 to 112. Using the same instrument, the media professional was also requested to check a Likert-type scale to indicate the perceived importance of each task. Anchors for this scale were "little or no importance" and "very important."

The activities were selected to provide a continuum of activities within instructional development. In Educational Technology: Definition and Glossary of Terms, instructional development is divided into six functions including Research-Theory, Design, Production, Evaluation-Selection, Logistics, and Utilization-Dissemination (AECT, 1977: 164). Each of the 28 activities is subsumed by at least one of the instructional development functions, with all six functions being represented by at least one activity.

The instrument was sent to several media professionals for their comment, utilized in a pilot study (Turner and Martin, 1978), and revised on the basis of responses and data received. A reliability coefficient of .94 was established using the split-half method.

Construct validity was investigated through the use of the factor analysis of the 28 activities. Six factors were formed, four of which accounted for 91.3 percent of the variance and included 25 of the 28 activities. A review of the activities included within each factor permitted the assignment as factor constructs of four of the six functions of instructional development delineated in the AECT definition. These factor constructs included Evaluation-Selection, Design, Production, and Logistics. The grouping of activities within the Media Professional Activities Checklist mirrored, to a large extent, the theoretical groupings proposed in the AECT definition of instructional development.

5. Principal's Perception Checklist. This instrument consisted of the 28 activities found on the Media Professional Activities Checklist. The respondent indicated, using a Likert-type scale, the importance of each task being performed by the media professional in the school. Possible scores ranged from 0 to 112.

Procedures

A packet containing the instruments and two cover letters was sent to the principal of each school in the sample. The initial cover letter was from the State Department of Education explaining the importance of the study and requesting cooperation. The second letter, which was from the researchers, introduced and provided the rationale for the study as well as supplying directions for the principal. The principal was requested to forward the instruments to be completed by the media professional and to see that all materials were returned. The media professional's materials were in a self-contained packet which was sealed before returning to the principal for mailing.

Data Collection

The responses from the instruments were tabulated with most of the responses being used in raw data form. A small percentage were translated into frequency counts. A total score was generated for the Empathy Scale, the Principal's Perception Checklist, and for each of the two parts of the Media Professional Activities Checklist. The other instruments were scored for each subsection.

Analysis of the Data

The variables for which dichotomous data were generated were analyzed by the use of chi-square tests of independence. To do this, the total score from the performance section of the Media Professional Activities Checklist was dichotomized around the mean into 0-66 and 67-112 categories. For the remaining independent variables, a simple correlation coefficient was generated against the independent variable.

The 28 activities of the Media Professional Activities Checklist were submitted to a factor analysis using the Varimax method of rotation. Factor scores were generated for each subject. Finally, the independent variables were run against the factor score variables in a stepwise multiple regression.

RESULTS AND ANALYSES

Two hundred and three questionnaires were returned. Of these, 187 were complete enough to be usable.

Dichotomous Data Analyses

Seven of the 40 variables yielded dichotomous data. These included grade levels of students served, student/media professional ratio, location of AV services in or out of the department, availability of a production area, released time for teachers to engage in extended planning, sex of, and degree(s) held by, media professionals. Of these variables, only the released time for teachers for extended planning yielded a significant difference at the .05 level. (See Table 1 for chi-square values.) Two of the remaining three variables were significant beyond the .1 level. These were the student/media professional ratio and availability of a production area.

TABLE 1

RESULTS OF CHI-SQUARE TESTS OF INDEPENDENCE

Variable	x^2	df
Grade levels served	4.716	7
Student/media professional ratio	15.20*	9
Location of AV services	0.38	2
Presence of AV production area	2.70*	1
Released time for teachers for extended planning	6.18**	1
Sex of media professionals	1.68	1
Type of degree (Bachelors, Masters, or Sixth-year)	1.02	5

*$p < .1$ **$p < .02$

As can be seen in Table 2, the significant difference generated by the teacher released time variable resulted in a greater number of instructional development activities reported by media professionals in schools where teachers were given time off during the year for planning. The presence or absence of a production area, as evidenced in Table 2, also resulted in a significant difference, although at the .1 level of significance, in the amount of instructional development reported. A greater proportion of the media professionals who had a separate production area reported that they performed instructional development than those without such facilities.

Results and Analyses of the Remaining Variables

A Pearson's r was calculated for each of the remaining variables to establish the relationship between these variables and the amount of instructional development reported. As can be seen in Table 3, four of the variables: hours assigned to the media center, number of student helpers, number of parent volunteers, and age of the media professional, were correlated negatively with the amount of instructional development activities reported. The correlations, however, are all low, with only the number of student helpers generating a significant relationship and that at the .1 level.

Of the 29 variables with positive relationships, 21 generated relationships significant at the .1 level or above. Variables significant at the .01 level or above included number of production equipment items, number of professional journals read, number of innovative teaching structures, all five of the competency variables, empathy, the principal's attitude toward the performance of instructional development by the media professional, and the media professional's attitude toward instructional development.

TABLE 2

CONTINGENCY TABLES FOR TEACHER RELEASED TIME AND AV PRODUCTION AREA VARIABLES

Teacher released time	Instructional Development		
	0-66	67-112	
Yes	6	17	23
No	69	54	123
	75	71	146

df 1

$x^2 = 6.18$

$p < .02$

AV Production Area	Instructional Development		
	0-66	67-112	
Yes	28	41	69
No	49	40	89
	77	81	158

df 1

$\underline{x}^2 = 2.71$

$\underline{p} < .1$

TABLE 3

MEANS AND STANDARD DEVIATIONS OF CONTINUOUS DATA INDEPENDENT VARIABLES AND r VALUES WITH REPORTED INSTRUCTIONAL DEVELOPMENT

Variable	Mean	SD	r with reported instructional development
Hours assigned in media center	6.65	1.32	-.0338
Time not assigned to supervisory tasks	1.40	1.74	.0437
Enrollment	691.35	410.28	.0403
Budget for print materials	1983.40	1660.41	.1081*
Budget for production supplies	249.24	411.00	.1884**
Budget for non-print commercial materials	612.44	858.00	.1901**
Total budget	2356.16	2456.66	.1769**
Number of services provided by district	1.48	1.35	.0643
Number of production equipment items	1.93	1.45	.3656****
Number of professional journals read	3.16	1.63	.2204***
Amount of teachers' daily planning time (hours)	.83	.42	.0624

*$p < .1$

**$p < .05$

***$p < .01$

****$p < .001$

Table 3 continued

Variable	Mean	SD	r with reported instructional development
Amount of paid help (hours daily)	1.09	1.64	.1133*
Number of student helpers/hour	1.87	2.17	-.1084*
Number of parent volunteers/hour	.05	.26	-.0824
Number of teaching structures	1.89	1.39	.1905***
Age of media professional	40.03	11.47	-.0960
Years of classroom experience	6.22	7.86	.0155
Years of library experience	8.53	6.22	.0680
Years at present position	6.56	5.37	.1029*
Years since coursework	3.71	3.94	.0333
Number of design courses	5.45	3.06	.1173*
Number of traditional courses (library)	4.26	1.20	.0895
Number of education courses	3.02	1.73	.1620**
Number of production courses	.52	.76	.1797**
Total number of courses	13.28	5.75	.1517**
Competency in design	22.74	7.03	.4031****
Competency in traditional library skills	17.42	2.87	.3218****

$^{*}p < .1$ $^{***}p < .01$

$^{**}p < .05$ $^{****}p < .001$

Table 3 continued

Variable	Mean	SD	r with reported instructional development
Competency in education	11.19	4.10	$.3674^{****}$
Competency in production	2.64	2.05	$.2285^{***}$
Total competency	53.64	12.83	$.4483^{****}$
Empathy	15.87	3.69	$.2224^{***}$
Principal's attitude toward media professional performance of instructional development	89.02	13.50	$.3559^{****}$
Media professional's attitude toward the performance of instructional development	90.18	17.27	$.4084^{****}$

$^{*}p < .1$

$^{**}p < .05$

$^{***}p < .01$

$^{****}p < .001$

Factor Analysis

A factor analysis of the Media Professional Activities Checklist utilizing the Varimax Method was undertaken. This analysis resulted in the formation of four factors including 25 of the 28 activities and accounting for 91.3 percent of the variance (See Table 4).

TABLE 4

FACTOR GROUPINGS OF INSTRUCTIONAL DEVELOPMENT ACTIVITIES

Factor 1	Evaluation-Selection
Verbally suggests titles of instructional materials to teachers	
Determines measurable objectives for media program	
Solicits feedback from teachers via forms and interviews about materials produced and purchased	
Validates the instructional materials purchased and produced by using observations of student behavior and matching these with instructional objectives	
Prepares lists of commercial materials in response to instructional objectives, learner characteristics, and other parameters	
Engages in research activities relative to the media center program	
Develops lists of titles of instructional materials in response to a specific topic given by a teacher	
Works with teachers in selecting materials to meet specific objectives	
Eigen Value 8.6965	Percentage of Variance 68.6

Factor	Design
Works with teachers in formulating instructional objectives	
Plans and discusses media center-related topics with teachers	
Participates in team-teaching activities	

TABLE 4 continued

Observes classroom work to coordinate media center activities with instructional programs

Participates in curriculum development and revision

Plans jointly with faculty members to coordinate media center activities with curriculum programs, units, and textbooks

Conducts workshops for teachers in evaluation and selection of materials

Acts as part of an instructional team in designing instructional programs

Develops long-range plans cooperatively with teachers and administration

Assists individual teachers in curriculum planning

Conducts workshops for teachers in use of materials and/or equipment

Eigen Value 1.3094 Percentage of Variance 10.3

Factor 3 Production

Designs and produces instructional materials in response to a general topic given by a teacher

Designs and produces instructional materials in response to instructional objectives, learner characteristics, and other parameters

Produces instructional materials in response to a specific design given by the teacher. This design specifies the type of medium and content to be included

Eigen Value .8954 Percentage of Variance 7.1

Factor 4 Logistics

Arranges and conducts sessions to preview and evaluate materials

Gives instructions to students in use of materials and/or equipment

Maintains a source of curriculum guides and professional materials for use in planning

Eigen Value .6717 Percentage of Variance 5.3

Constructs taken from the sub-functions of instructional development as defined in Educational Technology: Definition and Glossary of Terms (AECT, 1977: 164) were assigned to these four factors. These included Evaluation-Selection, Design, Production, and Logistics. Factor-scores based on the four factors were generated for each subject.

Multiple Regression

Four stepwise regressions were run with each factor utilized as a dependent variable associated with the independent variables. These runs were set to limit the independent variables utilized to those that contributed at least two percent additional variance.

The first factor was assigned the construct Evaluation-Selection. The independent variables utilized in the equation, in order entered, were total competency, hours assigned, number of design courses, principal's attitude, number of education courses, classroom experience, amount of time not assigned supervisory tasks, and competency in traditional library skills (see Table 5). Two of the variables, number of design courses and amount of classroom experience, related negatively to the Factor 1 variable. The eight variables utilized accounted for 54 percent of the variance.

The second factor was assigned the construct of Design. The independent variables utilized in the multiple regression equation were total competency, principal's attitude, total budget, amount of time not assigned supervisory tasks, competency in design area, number of education courses, attitude of media professional, and competency in education courses (see Table 6). All of the relationships between the independent variables utilized and the factor-score were positive. The resulting equation accounted for 68 percent of the variance.

TABLE 5

MULTIPLE REGRESSION SUMMARY TABLE FOR FACTOR 1 (EVALUATION-SELECTION)

Independent Variables	Multiple R	R Square	RSQ Change	Simple R
Total competency	.5176	.2697	.2697	.5167
Number of hours assigned to media center	.5665	.3209	.0530	.1792
Number of design courses	.6194	.3836	.0627	-.0268
Principal's attitude	.6627	.4392	.0555	.2972
Number of education courses	.6861	.4707	.0314	.1352
Classroom experience	.7033	.4947	.0239	-.0632
Amount of time not assigned supervisory duties	.7183	.5160	.0213	.0710
Competency in traditional library skills	.7387	.5456	.0295	.4456

TABLE 6

MULTIPLE REGRESSION TABLE FOR FACTOR 2 (DESIGN)

Independent Variables	Multiple R	R Square	RSQ Change	Simple R
Total competency	.6274	.3936	.3936	.6274
Principal's attitude	.6884	.4739	.0803	.4190
Total budget	.7222	.5216	.0476	.3024
Amount of time not assigned supervisory duties	.7482	.5599	.0383	.1664
Competency in design	.7757	.6081	.0419	.5290
Number of education courses	.7956	.6329	.0311	.0937
Media professional's attitude	.8142	.6629	.0299	.1959
Competency in education area	.8276	.6849	.0220	.5432

The construct of Production was assigned to the third factor. Six independent variables were utilized in the multiple regression equation to explain 46 percent of the variance. The variables included total competency, number of parent volunteers, enrollment, amount of time not assigned supervisory tasks, competency in traditional library skills, and competency in production (see Table 7). Two of the variables, number of parent volunteers and enrollment, were associated negatively with the factor-score.

TABLE 7

MULTIPLE REGRESSION SUMMARY TABLE
FOR FACTOR 3 (PRODUCTION)

Independent Variable	Multiple R	R Square	RSQ Change	Simple R
Competency total	.4852	.2354	.2354	.4852
Parent volunteers	.5930	.3516	.1162	-.3052
Enrollment	.6274	.3937	.0420	-.0134
Amount of time not assigned supervisory tasks	.6447	.4157	.0219	.1709
Competency in traditional library skills	.6612	.4372	.0214	.3046
Competency in production	.6791	.4655	.0283	.3287

The fourth and final factor was assigned the construct of Logistics. Ten independent variables were entered into the equation resulting in the explanation of 62 percent of the variance. These included competency in the education area, number of production equipment items, years of classroom experience, production budget, amount of student help, amount of paid help, number of traditional library courses, hours assigned to the media center, age, and years since last course work. Four of the variables were related negatively to the factor variable. These included years of classroom experience, amount of student help, age, and years since last coursework.

TABLE 8

MULTIPLE REGRESSION SUMMARY TABLE FOR FACTOR 4 (LOGISTICS)

Independent Variable	Multiple R	R Square	RSQ Change	Simple R
Competency in education	.4385	.1923	.1923	.4385
Number of production equipment items	.5734	.3288	.1365	.3781
Years of classroom experience	.6248	.3904	.0615	-.1504
Production budget	.6528	.4332	.0428	.3343
Student help	.6822	.4653	.0321	-.2449
Number of traditional courses	.7317	.5379	.0254	.1147
Number of hours assigned to media center	.7474	.5615	.0236	.2773
Age	.7596	.5858	.0243	-.0571
Years since last coursework	.7872	.6237	.0379	-.0424

CONCLUSIONS

The leap from relationship to causality is a critical one and can only be justified when the methodology is valid and the instrumentation is accurate. Because this study enters terra incognita, the leap requires a large assist from faith as well as from logic. Some distinct patterns, however, have evolved out of this study, and these warrant elaboration.

The variables that stand out from the others in consistency and in strength of positive relationship to the amount of reported instructional development are four competency components taken singly: design, traditional library skills, education, production, as well as the sum total of these components. An immediate tendency exists to be skeptical about a high degree of correlation between two measurements when the raw data were self-reported. There is evidence, however, that indicates that use of the reporting form for competencies is not necessarily subject to a distortion caused by an inordinately favorable reporting of competency and performance. In a pilot study (Turner and Martin, 1978), utilizing a population with relatively homogeneous competencies, the competency variables did not prove to be significantly related to the amount of reported instructional development.

The first conclusion, therefore, is that the reported competencies of the media professional in the areas of design, traditional library skills, education, and production have the strongest relationship with reported instructional development. This conclusion is, of course, not surprising as one's competency in an area, especially in a predominantly elective sector such as instructional development, should certainly have a strong bearing on whether activities are undertaken or not. It would be difficult to envision a strong instructional development component within a K-12 media program if the media professional lacked competency in any of the above four areas.[1]

---

[1] The reading of professional journals, certainly a method of increasing competence, is also positively related to instructional development. This variable was also significant in the pilot study where continuing independent study was the main method of differentiation of competency for the population utilized.

The second conclusion is that a positive attitude by both the principal and the media professional toward the inclusion of instructional development in the role of the media professional increases the amount of instructional development undertaken. Once again, this is not a startling conclusion, being consistent with the findings of several previous studies (Larsen, 1971; Leeper, 1975). Since the performance of instructional development is so often elective, done beyond the normal routine, and often requires a change in administrative factors, a positive attitude by the media professional as well as the principal would obviously be beneficial.[1]

The third conclusion is based on the low but significant relationships of all four budget variables, as well as the moderate relationship evidenced with the number of production equipment items present. This conclusion is that the performance of instructional development requires a minimum financial support base. Since instructional development usually requires alternative materials and delivery systems, as well as the production of materials, this minimum base requirement is to be expected.

The fourth conclusion concerns the empathy variable. While the relationship between the empathy score and reported instructional development is low, it is significant and provides the basis for the conclusion that empathy is a positive attribute of the personality of the instructional developer at the K-12 level.

---

[1] The availability of released time for teachers to engage in extended planning, a variable which could also be influenced by the principal's attitude, resulted in a significant difference in the chi-square analysis and lends weight to this conclusion.

The fifth conclusion is actually a corollary of the second in that the principal usually makes the general schedule, and the media professional generates the specific details. This conclusion is that in order to perform instructional development activities, the media professional must be assigned to the center with sufficient time to perform logistical support activities. More importantly, the media professional must be allowed time where he/she is not involved in direct supervisory activities in order to perform evaluation, selection, design, and production activities.

The sixth, and final, conclusion is based on the negative relationship found in this study and in the pilot study of the parent volunteer and student help variables to the reported amount of instructional development. This conclusion is that volunteer assistance in the form of either students or parents does not appear to contribute to the amount of instructional development undertaken. This rather surprising conclusion may be explained by the amount of time that is required to organize and administer volunteers. Rather than freeing the media professional to perform instructional development activities, energies appear to be channeled into supervisory activities.

RECOMMENDATIONS

Based upon the conclusions of this study, the following recommendations are offered:

1. The graduates of programs preparing media professionals for the K-12 level must be provided a reasonable level of competency in the areas of traditional librarianship, traditional education, design (learner analysis, specification of objectives, systems design, validation, etc.), and production. The decision as to whether a graduate has such competencies should not be based solely on the number of courses completed, but should be based on competency testing.[1]

2. The faculty of the programs that prepare K-12 media professionals should initiate programs of in-service training to upgrade the competencies of media professionals currently in the field.

3. Efforts must be made to foster a positive attitude in the building principal. This can be done through seminars at the meetings of school administrators' associations and through publishing in the literature that reaches these administrators.

4. Local, state, and national organizations must continue searching for funds to provide an adequate budget on which to base an instructional development program.

5. Further research should take place into the importance of an empathetic personality for the performance of instructional development and the efficacy of empathy training in affecting personality. Continued research should also take place in the area of empathy training.

6. The media professional should be provided with at least one period a day in which no supervisory tasks are required. Such a period would be analogous to the teachers' "planning" period.

---

[1] It will be remembered that the number of courses had a much weaker relationship with instructional development than the competency variables.

7. Finally, K-12 media professionals should be provided training in personnel management to allow them to manage volunteers more effectively.

Continued work on the development of instruments to measure the amount and quality of instructional development performed at the K-12 level needs to be undertaken. Further studies should be performed which utilize observational as well as questionnaire methodologies. With the emergence of information pertaining to the characteristics of the successful instructional developer, perhaps the future will provide us with the realization of the promise held forth in the literature--each media professional working alongside other educators toward the facilitation of the curriculum.

REFERENCES

American Library Association. Task Analysis Survey Instrument. Chicago: ALA, 1969.

________. Evaluation of Alternative Curricula. Chicago: ALA, 1975.

________ and Association for Education Communications and Technology. Media Programs: District and School. Chicago: ALA, 1975.

Association for Educational Communications and Technology. Educational Technology: Definition and Glossary of Terms. Washington: AECT, 1977.

Brown, James W., Norberg, Kenneth D., and Srygley, Sara K. Administering Educational Media. 2nd ed. New York: McGraw-Hill, 1972.

Daniel, Evelyn Hope. The Organizational Position and School Media Centers: An Analysis of the Role of the School Library and the School Librarian (Doctoral Dissertation, University of Maryland, 1974). Ann Arbor, Mich.: University Microfilms, 1974, No. 74-29, 065.

Davies, Ruth Ann. The School Library: A Force for Educational Excellence. New York: R. R. Bowker, 1969.

Erickson, Carlton W. H. Administering Instructional Media Programs. New York: Macmillan, 1968.

Grief, E. B., and Hogan, R. "The Theory and Measurement of Empathy." Journal of Counseling Psychology, 1973, 20, 280-84.

Hogan, R. "Moral Conduct and Moral Character: A Psychological Perspective." Psychological Bulletin, 1973, 79, 217-32.

________ and Dickstein, E. "A Measure of Moral Values." Journal of Counseling and Clinical Psychology, 1972, 39, 210-214.

________ and Mankin, D. "Determinants of Interpersonal Attractions: A Clarification." Psychological Reports, 1970, 26, 335-38.

Jetter, Margaret A. The Roles of the School Media Specialist in the Future: A Delphi Study (Doctoral Dissertation, Michigan State University, 1972). Ann Arbor, Mich.: University Microfilms, 1973, No. 73-12, 746.

Kerr, S. T. "Are There Instructional Developers in the Schools? A Sociological Look at the Development of a Profession." AV Communications Review, 1977, 25, 243-68.

Kurtines, W., and Hogan R. "Sources of Conformity in Unsocialized College Students." Journal of Abnormal Psychology, 1972, 80, 49-51.

Larsen, John A. *The Role of the Media Specialist as Perceived by Himself and His Administrator in the Secondary Schools of Utah* (Doctoral Dissertation, University of Utah, 1971). Ann Arbor, Mich.: University Microfilms, 1971, No. 71-24, 932.

Leeper, Dennis P. *A Comparative Study of Open-Space and Self Contained Elementary School Library Media Centers* (Doctoral Dissertation, University of Colorado, 1975). Ann Arbor, Mich.: University Microfilms, 1976, No. 76-3921.

Maudaus, James M. *Curriculum Involvement, Teaching Structures, and Personality Factors of Libraries in School Media Programs* (Doctoral Dissertation, University of Texas at Austin, 1974). Ann Arbor, Mich.: University Microfilms, 1975, No. 75-4965.

Prostano, Emanuel and Prostano, Joyce. *The School Library Media Center.* Littleton, Colorado: Libraries Unlimited, 1971.

Schulzetenberge, Anthony C. *Interests and Background Variable Characterizing Secondary School Librarians Who Work With Teachers in Curriculum Development and Improvement of Instruction* (Doctoral Dissertation, University of North Dakota, 1970). Ann Arbor, Mich.: University Microfilms, 1972, No. 72-16, 372.

Turner, Philip M., and Martin, Nina N. *Factors Affecting Instructional Development Activities of Selected K-12 Media Professionals.* University, Alabama: The University of Alabama, 1978. (ERIC Document Reproduction Service No. ED 151 012).

PART III:

MEDIA CHARACTERISTICS RESEARCH

The Effects of Picture Type and Picture Location on Comprehension

Philip J. Brody
Alice Legenza
University of Kansas

A pre-publication draft of a paper presented at the annual convention of the Association for Educational Communications and Technology, New Orleans, March, 1979.

Abstract

The Effects of Picture Type and Picture Location on Comprehension

Philip J. Brody & Alice Legenza

University of Kansas

Using mathemagenic-based research as a guide, this study attempted to determine if the location of a picture (pre- or post-reading passage) or the type of picture (overview or specific incident) could affect reading comprehension. Ninety-two college students were randomly assigned to one of the cells in a two-by-two factorial design and required to read a text-like passage of approximately 1200 words. Results indicated that pictures placed after the passage increased incidental learning more than pictures placed before the passage.

Mathemagenics have been studied since the 1960's, when Rothkopf (1966) defined the term to refer to a broad array of behaviors which increase learning. Although mathemagenics have provided a theoretical base for systematic research in some areas (Faw and Walker, 1976), it has largely been neglected by researchers in the field of media.

While the mathemagenic effects of various instructional techniques such as objectives and advance organizers have been studied, the most widely and systematically studied area has been in the use of questions. Typically (e.g., Frase, 1967), in these studies, students are asked to read passages in which questions have been inserted either before or after the passage (pre- and post-questions). Students are then given a multiple choice exam over the material. The results of these studies generally suggest that both pre- and post-questions facilitate learning that is specific to the questions asked while post-questions have greater effect on learning that is not directly related, or is incidental, to the questions.

Research which has attempted to relate mathemagenics and media has generally incorporated pre- and post-questions into various presentational formats (e.g., Lavin, 1972). Dayton (1977), while describing how mathemagenics can be used in the design of slide/tape presentations, neglects to discuss the important implcations of the mathemagenic hypothesis on the non-verbal aspects of the presentation.

An area of media research which would seem to benefit from mathemagenics based research would be the relationship between pictures and reading comprehension; some researchers (e.g., Samuels, 1970; Concannon, 1975) suggest that pictures may interfere with comprehension, while others (e.g., Pressley, 1977; Levin and Lesgold, 1978) suggest that they can be beneficial. Much of the research that has been done on the effects of pictures on comprehension suffers from one of several confounding variables. Peeck (1974), for example, utilized materials which were not representative of those found in schools. Other research has either made little attempt to define the pictorial attributes which were being studied or has done so in terms of the well-worn realism continuum (Dwyer, 1972). How then can mathemagenics serve as a guide for research in this area?

It would seem, first of all, that the location of pictures within a reading passage could be of importance. It needs to be established if pre- and/or post-pictures have the same mathemagenic functions as pre- and post-questions. Mathemagenic research would seem to suggest that pictures placed after a reading passage would have a greater effect on "incidental" learning (that which cannot be learned directly from viewing the picture) than would pictures placed before the passage; similarly, pictures which provide a broad overview should be superior to those which provide a narrow focus. Thus, the purpose of this study was to 1) determine the extent to which picture location and picture type serve mathemagenic functions and 2) to determine the extent to which these variables affect reading comprehension.

Method

Subjects

Subjects were 92 undergraduate students enrolled in education methods courses.

Materials

A passage of approximately 1200 words about the square of Marrakesh was constructed from a variety of books and encyclopedias. The main theme of the passage was the daily activities of the square, although half the article discussed history, geography, and other related background information.

Two black-and-white pictures concerning the square were also reproduced; the first picture, a summary picture, was an aerial view of the square, showing the square and its surroundings. The second picture showed a close-up of a specific incident taking place in the square, a snake charmer with his snakes.

Twenty-four multiple choice questions were developed over the passage. Half of the questions were over the background information, while the other half were over the square activities; but all questions assessed incidental, rather than intentional, learnings.

Procedure

Subjects were divided into 4 groups, with 23 subjects per cell. The groups were divided on two levels of picture viewing - 1) before reading the passage, and 2) after reading the entire passage. These groups were further subdivided so that half of the subjects within each level of the picture viewing saw the summary picture while the other half of the subjects saw the specific incident picture.

On the day of the experiment each subject was given a consent form to sign which stated that the investigators were doing research on media and reading. Subjects were randomly given booklets containing one of the pictures in one of the two locations and told to "study each page of the booklet for as long as you want but you may not turn back to any page after you have turned to the next page. When you are done, raise your hand and you'll be given a short test to take."

Results and Discussion

The means and standard deviations of correct answers on the comprehension test are presented in Table 1.

Insert Table 1 about here

As indicated in Table 2, picture location effects were significant at the

Insert Table 2 about here

.05 level (F (1,88) = 4.22) for the square activity questions, with post-pictures producing higher comprehension scores. For the square activity questions, the effects of picture type was not significant. Neither picture location nor picture type produced significant results for "background information questions" (see Table 3).

Insert Table 3 about here

The results of this study add preliminary support to the idea that pictorial attributes can serve mathemagenic functions. As with "inserted questions" reasearch, this study found location to be a significant instructional variable since pictures placed after a reading passage increased incidental learning more than pictures placed before the reading passage. However, the lack of significant results related to the "picture type" variable indicates that not all pictorial attributes affect mathemagenic behaviors.

These mixed results call for additional research to more precisely determine the possible mathemagenic effects of pictorial attributes. Most importantly, the study needs to be replicated with different audiences, particularly with younger subjects and/or those with different comprehension skills than the college students utilized. Although it seems logical to assume that the results should be replicated or even magnified when using younger subjects, results of research with both reading behaviors (Knafle and Legenza, 1978) and picture utilization behaviors

(Pressley, 1977) differ with age of subjects.

Future research must also focus upon those characteristics which will optimize the mathemagenic effects of pictorial attributes. For example, the optimum length of the passage between pictures needs to be determined, as does the complexity of both the reading passage and picture. Other pictorial attributes that affect mathemagenic behaviors also need to be identified and studied. These additional attributes should be applicable to a wide range of educational media, rather than being applicable to pictures only (Salomon, 1978).

To the extent that the research being reported is replicated and expanded in the directions indicated, instructional design specialists as well as all other classroom teachers will possess an instructional technique which can provide substantial benefits.

Table 1

Means and Standard Deviations on

Correct Answers on the Comprehension Test

Treatment	Square Activity		Background	
	$\overline{X}$	sd	$\overline{X}$	sd
Pre snake charmer	6.0	2.28	3.35	1.46
Pre square activity	6.39	1.95	3.61	1.64
Post snake charmer	7.04	1.94	3.17	1.99
Post square activity	7.13	1.94	3.91	2.06

Table 2

Analysis of Variance for "Square Activity" Questions

Source	SS	df	MS	F
Location	18.27	1	18.27	4.22*
Type	1.31	1	1.31	.30
LxT	.53	1	.53	.13
Within	381.05	88	4.33	
Total	401.16	91		

*p < .05

Table 3

Analysis of Variance for Background Information Questions

Source	SS	df	MS	F
Location	.10	1	.1	0
Type	5.75	1	5.75	1.69
LxT	1.31	1	1.31	.38
Within	299.83	88	3.41	
Total	306.99	91		

References

Concannon, S. J. Illustrations in books for children: Review of Research. Reading Teacher, 1975, 29, 254-256.

Dayton, D. K. Inserted post-questions and learning from slide-tape presentations--implications of the mathemagenic hypothesis. AV Communication Review, 1977, 25, 125-146.

Dwyer, F. M. A guide for improving visualized instruction. State College, Pa.: Learning Services, 1972.

Faw, H. W., and Waller, T. G. Mathemagenic behaviors and efficiency in learning from prose. Review of Educational Research, 1976, 46, 691-720.

Frase, L.T. Learning from prose material: Length of passage, knowledge of results, and position of questions. Journal of Educational Psychology, 1967, 58, 266-272.

Knafle, J. D. and Legenza, A. External generalizability of inquiry involving artificual orthography. American Educational Research Journal, 1978, 15, 331-347.

Lavin, M. H. Facilitative effects of mathemagenics narration in auditape-slide programs in pathology. Doctoral dissertation, University of Iowa. 1971. Dissertation Abstracts International, 1972, 32, 5108A. (University Microfilms No. 72-8279).

Levin, J. R. and Lesgold, A. M. On pictures and prose. Educational Communications and Technology, 1978, 26, 233-243.

Peeck, J. Retention of pictorial and verbal content of a text with illustrations. Journal of Educational Psychology, 1974, 66, 880-888.

Pressley, M. Imagery and children's learning: Putting the picture in the developmental perspective. Review of Educational Research, 1977, 77, 585-622.

Rothkopf, E. Z. Learning from written instructional materials: An exploration of the control of inspection behavior by testlike events. American Educational Research Journal, 1966, 3, 241-249

Salomon, G. On the future of media research. Educational Communication and Technology, 1978, 26, 37-46.

Samuels, S. J. Effects of pictures on learning to read, comprehension, and attitudes. Review of Educational Research, 1970, 40, 397-407.

RECALL AND RETRIEVAL
FROM MAPPED AND PROGRAMMED TEXT

David H. Jonassen, Ed.D.
Assistant Professor
School of Education
University of North Carolina at Greensboro
Greensboro, North Carolina 27412

Running Head: MAP/PI

Abstract

Two methods of structuring textual material, inormation mapping and programmed instruction, were compared on the basis of their ability to facilitate recall and subsequent retrieval of information. Both methods produced very significant gain scores, however no differential effects occurred. Retrieval of information from the text, a task independent of recall behavior and based upon different textual search strategies, benefitted significantly from the structural characteristics of the mapping technique. A conceptual basis for the issue being studied was suggested: a comparison of systematically inserted adjunct questions and explicit content structuring for eliciting mathemagenic activities. Suggestions of appropriate literatures such as structural linguistics are somewhat conjectural because of the absence of any research base for mapping.

Introduction

A reversionary trend toward print, textual media for instruction is based upon the increasing importance of cost effectiveness as a selection criterion. Audiovisual media, in varying degrees of sophistication, have failed to produce the panaceas predicted with their development. Coupled with the continuing preeminance of print (verbal mode) instructional media, the need to explore the potential of text for information transmission and retrieval has never been greater.

One of the most promising print methodologies, information mapping[1] (MAP), will be compared with another prominent method for structuring prose, programmed instruction (PI). This study constitutes a seminal investigation, aimed at documenting main effects (evaluative comparison) and suggesting some theoretical bases for considering these instructional methods. As such, it should raise more questions than it answers, however, this seems appropriate considering the embryonic state of information mapping research.

Programmed Instruction

Programming textual materials normally consists of limiting the size of information chunks, requiring overt responses to the information presented, providing feedback (reinforcement), and allowing for individual pacing. Evaluative research or programm instruction abounds, the reviews (Lange, 1972; Schramm, 1964) generally indicating a slight learning advantage and a more significant time savings for PI when compared with traditional instruction. In analyzing the contributions of the aforementioned characteristics of PI, Hartley (1974) concluded that none were essential to the success of any program, that removing one or more of these techniques has little if any effect on the learning results.

---

[1] Information Mapping^{TM} is a trademark of Information Resources, Inc.

The intention of this study is to identify a common basis for comparing media, not replicating research on programming attributes. The programmed version of text prepared for this study could alternatively be defined as the systematic insertion of adjunct questions (Rothkopf, 1972) into textual material, as multiple choice questions were consistently inserted after chunks of prose, with knowledge of results presented immediately. While the programmed material compared with mapping in this study includes all of the characteristics of programming cited above, it could easily be sumsumed by another rubric. The naming of the technique is not as important as the underlying cognitive processes.

The issue of importance is the degree to which consistent overt responding to textual material with subsequent feedback will evoke recall behavior. Several studies have supported the facilitative effect of inserted post-questions, especially for intentional learning, on immediate posttests (Frase, 1968; Rothkopf, 1972; Rothkopf & Bibiscos, 1967). Learning can be expected to improve as a function of the programming technique employed in this study.

Information Mapping

Information mapping is a technique for developing textual materials developed by Robert Horn almost a decade ago (Horn, Nicol, and Kleinman, 1969). It consists of a specific set of rules and procedures for classifying, organizing, and presenting information. The primary characteristics include chunking information into units and presenting it in labelled blocks, the marginal labelling providing an intrinsic indexing system. With the presentation format remaining consistent, these information blocks are sequenced congruously with the nature of the information being presented into superordinate structures or maps (structure, procedure, process, fact, concept and principal types). The structural characteristics externally organize information, with hypothesized facilitation of information coding and retrieval.

As an instructional methodology, information mapping is only beginning to emerge. Used as format for presenting formal reports (Hartjen, 1977), it has been recommended as a valuable technique for instruction (Thiagaragan, 1977), industrial training (Horn, 1975), computer assisted instruction (Horn, Nicol, Roman, & Razar, 1971) and task analysis (Shroeder, 1975). Surprisingly little literature has chronicled its development as an important medium for industrial communication. Even less empirical analysis exists, although research is currently in progress, especially in England. The two most prominent studies failed to document any learning benefits from mapping when comapred with prose (Horn, Nicol, & Kleinman, 1969) or with computer assisted instruction (Geisert, 1970), however attitudes toward mapped texts were superior in both. However, neither study provides any theoretical or empirical support for the claims of its developer.

As a seminal effort, this study focused on the potential of some of the major characteristics of programmed and mapped texts for eliciting recall and as a medium for information retrieval. Conceptual foundations will be considered later.

Recall. Since no analysis or manipulation of mapping variables has been done, inferences or predictions about mapping characteristics must derive from other literatures, specifically that related to the semantic structure of prose, an issue which has received considerable attention. Prose passages normally contain an implicit organizational structure of content. Recall of ideas becomes a function of the height in the content structure of a given idea, i.e., ideas high in the structure are better recalled (Meyer & McConkie, 1973). This engendered the concept of staging (Grimes, 1972), the dimension of prose structure that determines the prominence of ideas within prose. Clements (1976) found that prose segments at high staging levels were recalled better.

This type of research convinced Meyer (1975) that she could "signal" this structure to the learner by adding non-content signals ("There are two approaches. One is . . . ") that would emphasize aspects of the content or its structure. Improved recall resulted, however additional research is needed. This type of research sets a precedent for the extrinsic reporting of prose structure, as occurs in information mapping, where types and levels of information are labelled allowing the learner to relate aspects of content together using the techniques of blocking and labelling.

The blocking of information produces the effect of associating linguistic elements together in a holistic fashion. Information enclosed in a block becomes a unit, permitting comparison of within-block semantic similarities to differences in elements presented in other blocks (Gropper, 1970). This chunking of information can also be reasonably expected to facilitate encoding of locative information about passages into memory. Christie and Just (1976) corroborated earlier results in concluding that individuals accurately remember the location of information in text. These incidental locators can act as spatial attributes of memory (Underwood, 1969) which function to discriminate one memory from another, thereby facilitating retrieval of the associated information from memory. The extrinsic blocking of information can only be expected to strengthen these spatial cues, facilitating recall of information intext.

Predictions about labelling must be derived from the literature on headings. Functioning as advanced organizers, headings provide conceptual cues about the ensuing information in a passage, providing an "explicit structure which assists the reader in integrating the information as he reads" (Wright, 1977, p. 96). Recall has been facilitated by subsuming information under major subheadings (Dahl, 1973) or by presenting one line titles immediately prior to the passage (Dooling and Lachman, 1971). Wright (1977) recommended differentiating these headings from the body of the text, either spatially or typography, as is done in information mapping.

Retrieval. Since information mapping has been recommended as an effective tool, consideration of the information retrieval characteristics seem important. However, the literature related to information retrieval is concerned either with memory processes or automated retrieval systems modelled after the former. Techniques for facilitating the retrieval of information from text material, such as those suggested by information mapping (table of contents, predictable textual format, consistent typography, marginal labels, and local indexing) operate in an atheoretical context, so that any successes forthcoming must be explained by other literature. Searching is generally agreed to be aided by section headings presented in the text and in a corresponding table of contents (Wright, 1977), but how these and other techniques work is uncertain.

Textual searching becomes a matter of associating locative information (spatial cues) stored in memory with their referent as it appears in a passage. Information mapping provides explicit cues (labels) which may be encoded as locative information in memory. Information retrieval requires associating only the locative cue with the label, thereby reducing memory load. At the very least, labels may be used as a convenient tag for searching text.

Some implications of blocking also seem obvious. Christie and Just (1976) hypothesized that "locative information could provide an index to the spatial distribution of information in the passage" (p. 706). Since retrieval of information from a passage, measured by eye fixations, was more efficient for organized passages, explicitly chunking information into units can only be inferred to provide stronger or more easily differentiable locative cues. Visual scanning should be accomplished more easily, thereby increasing textual retrieval capabilities.

Finally, the consistent format for sequencing information suggests a search strategy for locating information. Search rules are suggested by the materials because the learner knows conceptually where a particular attribute is located in relation to other material. This explicit sequencing should provide the basis for a consistent search procedure.

Method

Participants

Forty-one seniors and graduate students enrolled in two sections of an introductory media course were assigned to treatment. Attrition accounted for unequal N's (MAP = 22, PI = 19). Verbal ability level and subject background were assumed equivalent as no significant differences between groups in pretest performance occurred.

Procedures

All subjects were pretested and assigned to treatment. One week following the pretest, the self-study materials were presented to the participants, who were allowed to complete them at their own pace. Immediately following completion of the text, each individual turned in the material and requested the exam (same as pretest), a 45-item, four-option multiple choice exam. Test items were extracted verbatim from the text. The tests, in addition to the texts, were reviewed for content validity by three subject matter experts. Reliability of the instrument was established by the Spearman-Brown split-half technique ($r = .83$).

Two weeks following treatment all subjects were administered the retrieval examination, an open-book task consisting of 41 sentence-completion questions that required reference to the textual material. A total of one hour was allowed for this exam for all participants. To insure retrieval from the text occurred rather than recall behavior, subjects were required to indicate the page number in the text that contained the answer.

A pretest - posttest design was employed in this study, with each treatment group serving as the control for the other group. In subsequent studies a control group (narrative prose test) will be employed. The delayed test was not intended to assess the same dimension of cognitive functioning as the recall (pretest-posttest) task, so those results were analyzed separately.

Results

A 2x2 analysis of variance (ANOVA) (test X method) was calculated for the data presented in Table 1. A significant main effect for treatment, F (1, 39) - 340.92, p <.001, was experienced by both groups, though

---

Insert Table 1 about here

---

no differential effect of treatment (group X test) occurred, as gain scores were fairly equivalent. Since no significant differences in pretest scores and posttest scores occurred, it can be concluded that both the MAP treatment and the PI treatment produced similar amounts of learning.

An F test for homogeneity of variance in the retrieval test scores indicated equivalence (F (21, 18) = 1.11, $p > .05$) so that a t test for independent means was conducted. A significant difference (t (21, 18) = 4.14, $p < .001$) between the scores of the treatment groups occurred, indicating that the MAP group performed substantially better than the programmed group on the retrieval task. It is obvious from the data that the retrieval task was different and independent from the earlier recall task. The relationship between posttest scores and retrieval scores was extremely low and insignificant (r = .15) and the variances for the retrieval task were substantially higher. This indicates that this task was not facilitated by earlier learning (recall). In an attempt to control for the effects of prior learning, an analysis of covariance was conducted, indicating that recall (prior learning) was not a significant covariate with retrieval, confirming the independent nature of the retrieval task and the superiority of the MAP treatment for facilitating it.

Discussion

Recall

Significant learning (recall) resulted from both the PI and MAP techniques, indicating that for initial learning, both are adequate as presentational media. Without a control group treatment (narrative prose), the extent of recall facilitation resulting from their respective characteristics remains undocumented. It can only be concluded they produce equivalent amounts of learning. While initially disappointing, this equivalence in resultant learning provides a meaningful empirical context for suggesting a conceptual basis for their comparison.

The advantages of mapping are obviously structural, based on the premise that by breaking down information into small chunks and explicitly structuring the context, learning will be improved. Frase (1969) provides empirical support concluding that recall is enhanced by informing the reader of the structure of a passage prior to reading the material. By clearly defining the function of these information chunks and their conceptual relationship to other chunks, you are explicitly organizing material for memory. Like advanced organizers (Ausubel, 1968), mapping provides conceptual tags which function as an explicit "ideational scaffolding" for assimilating information. McConkie (1978) suggests that a complete understanding of the effects of prose structure on recall probably necessitates looking at some structural linguistic literature (Grimes, 1972; Meyer, 1975). For purposes of this discussion, suffice it to say that the important intrinsic characteristic of mapping is the structuring of content in textual passages. A meaningful context for comparing mapping and programming is the mathemagenic hypothesis.

The concept of mathemagenic activities (Rothkopf, 1970) should be familiar to most researchers. Meaningful prose learning has been consistently enhanced by the elicitation of mathemagenic activities in the learner. the covert mediational behaviors that produce learning, that determine "*the nature of the effective stimuli in experimental or instructional situations*" (Rothkopf, 1970, p. 326, emphasis his). Several techniques have been reviewed as capable of manipulating learners cognitive activities during acquisition, such as advance organizers, response modes (underlining, notetaking), objectives, and inserted questions (Faw & Waller, 1976). This list should include explicit content structuring that is provided by information mapping procedures. It can be tentatively concluded that this technique directly affects Class III mathemagenic activities, those concerned with visual scanning, translation into meaningful constructs, and internal cognitive processing (discrimination, differentiation) of prose (Rothkopf, 1970). It seems obvious that ordering and prearranging text will affect reading as do the other techniques suggested earlier. So the issue actually addressed in this study concerns the type and extent of mathemagenic activities elicited by consistently inserted questions vis a vis explicit content structuring. At the processing stage of Class III behavior, some form of organization of material inevitably occurs that will facilitate memory coding. The structuring implicit in the mapping technique could be hypothesized to supplant these mathemagenic activities (Ausburn & Ausburn, 1977). At any rate, it seems logical that mathemagenic activity is a meaningful context for comparing and analyzing the effects of these instructional methods on learning (recall).

Retrieval

The structural characteristics of mapping provide a clear advantage for retrieving information from textual materials. Some combination of the marginal labels, consistent format, and information blocking significantly facilitated the visual search procedure. In order to understand the contributions of each, these variables will have to be manipulated. It should be obvious from the previous discussion of mathemagenic activities that an alternative construct is needed to adequately deal with the retrieval process, since learning is not the issue of concern. Rather we need to be concerned with those textual characteristics that will maximally facilitate the search for specific information. A review of the reference and indexing literature should be combined with an analysis of automated information retrieval systems filtered through a visual information processing perspective, such as that suggested by Clark, Carpenter, and Just (1973). The retrieval issue is important and distinctive enough to be divorced from consideration of the learning potential of mapping, since different assumptions and analyses will be required for its proper consideration.

Summary

The potential of information mapping as an instructional technique seems obvious, however the virtual absence of any research base relegates mapping to the "potential" status. While this study was methodologically deficient (absence of control group, study time factor, prior experience, or concern with relevant learner aptitudes), it was intended only to globally address the issues and assumptions of mapping. The need for additional research to consider the level and type of learning and interactions with the characteristics of mapping and programming is unlimited. Additional research can be expected to clarify the usable parameters of both methods of instruction as well as shedding another glimmer of light on how learning occurs.

References

Ausburn, L. J. & Ausburn, F. B. A supplantation model for instructional design: A preliminary investigation. Paper presented at annual meeting of Association for Educational Communications and Technology, Miami, Fla., April, 1977.

Ausubel, D. P. Educational psychology: A cognitive view. New York: Holt, Rinehart & Winston, 1968.

Christie, J. M. & Just, M. A. Remembering the location and content of sentences in a prose passage. Journal of Educational Psychology, 1976, 68, 702-710.

Clark, H. H., Carpenter, P.A. & Just, M.A. On the meeting of semantics and perception. In W. G. Chase (Ed.), Visual Information Processing. New York: Academic, 1973.

Clements, P. The effects of staging on recall from prose. Unpublished doctoral dissertation, Cornell University, 1976.

Dahl, P. R. Memory for prose materials. Unpublished doctroal dissertation, University of Minnesota, 1973.

Dooling, D. J. & Lachman, R. Effects of comprehension on retention of prose. Journal of Experimental Psychology, 1971, 88, 216-222.

Falk, L.M. Retrieval of information from self-study texts. Manuscript in preparation, 1979.

Faw, H. W. & Waller, T. G. Mathemagenic behaviors and efficiency in learning from prose materials: Review, critique, and recommendations. Review of Educational Research, 1976, 46, 691-720.

Frase, L. T. Paragraph organization of written materials: The influence of conceptual clustering upon the level and organization of recall. Journal of Educational Psychology, 1969, 60, 394-401.

Geisert, P. A. A comparison of the effects of information mapped learning materials and traditional materials on the learning of concepts via the printed page and computer cathode ray tube. (Tech Memo No. 24). Washington: Office of Naval Research, 1970. (ED 046 253).

Gropper, G. L. The design of stimulus materials in response-oriented programs. AV Communication Review, 1970, 18, 129-159.

Hartjen, R. N. Compressed media: A new dimension of compressed speech. Paper presented at the annual meeting of Association for Educational Communications and Technology, Miami, Fla., April, 1977.

Hartley, J. Programmed Instruction 1954 - 1974: A review. Programmed Learning and Educational Technology, 1974, 11, Horn, R. E. More about information mapping. *Training*, 1975, *12* (2), 36-38.

Horn, R. E. *How to write information mapping*. Lexington, Mass.: Information Resources, 1976.

Norn, R. E., Nicol, E. H., & Kleinman, J. C. *Information mapping for learning and reference* (ESP - TR - 69 - 296). Bedford, Mass.: Command Systems Division, Electronic Systems Division, Air Force Systems Command, U.S. Air Force, 1969. (ED 042 322).

Horn, R. E., Nicol, E. H., Roman, R. A., & Razar, M. P. *Information mapping for computer-based learning and reference*. Cambridge, Mass.: Information Resources, 1971.

Lange, P. C. What's the score on programmed instruction. *Today's Education*, 1972, *61* (2), 59.

McConkie, G. W. Learning from text. In L. S. Shulman (Ed.), *Review of research in education*, 1977 (Vol. 5). Itasca, Illinois: Peacock, 1978.

Meyer, B. J. F. *The organization of prose and its effects in memory*. Amsterdam, The Netherlands: North-Holland, 1975.

Meyer, B. J. F., & McConkie, G. W. What is recalled after hearing a passage? *Journal of Educational Psychology*, 1973, *65*, 109-117.

Rothkopf, E. Z. The concept of mathemagenic activities. *Review of Educational Research*, 1970, *40*, 325-336.

Rothkopf, E. Z. Variable adjunct question schedules, interpersonal interaction, and incidental learning from written material. *Journal of Educational Psychology*, 1972, *63*, 87-92.

Rothkopf, E. Z. & Bibiscos, E. E. Selective facilitation effects of interspersed questions on learning from written materials. *Journal of Educational Psychology*, 1967, *58*, 56-61.

Schramm, W. *The research on programmed instruction: An annotated bibliography*. Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1964.

Shroeder, P. E. *Proceedings of a symposium on task analysis/task inventories*. Columbus, Ohio: Ohio State University Center for Vocational Education, 1975.

Underwood, B. J. Attributes of memory. *Psychological Review*, 1969, *76* 559-573.

Wright, P. Presenting technical information: A survey of research findings. *Instructional Science*, 1977, *6*, 93-134.

Table 1

Means and Deviations for Treatment Groups by Test

		Map Group	Programmed Instruction Group
Pretest (Recall)	$\bar{X}$	28.18	27.84
	SD	3.51	5.37
Posttest (Recall)	$\bar{X}$	38.86	37.53
	SD	2.87	3.37
Retrieval Test	$\bar{X}$	33.90	22.94
	SD	8.16	7.76
N		22	19

The Instructional Effectiveness of Integrating Abstract and Realistic Visualization

John H. Joseph
Assistant Professor of Education
Assistant to the Dean for Instruction
The Pennsylvania State University
The Capitol Campus

The Instructional Effectiveness of Integrating Abstract and Realistic Visualization

John H. Joseph
The Pennsylvania State University

The relative effectiveness of abstract and realistic visualization is debated in the literature and by practitioners. The study examined the instructional effectiveness of integrating abstract and realistic visualization for self-paced and externally-paced instruction involving five types of instructional tasks and students of different general ability.

The results suggest that designers of visualized instruction may expect to enhance its effectiveness under general conditions by inclusion of realistic visualization. However, the inclusion of abstract visualization or the integration of abstract and realistic visualization should be based on a consideration of pacing, general ability, and the type of instructional objective.

The Instructional Effectiveness of Integrating Abstract and Realistic Visualization

John H. Joseph

The Pennsylvania State University

Paper presented at the 1979 Annual Convention
Association for Educational Communications and Technology
Research and Theory Division
New Orleans, Louisiana March 7, 1979

Purpose

This study examined the question of whether the integration of abstract and realistic visualization can improve the effectiveness of visualized instruction. The question is derived from the perennial issue of the relative worth of realistic and abstract visuals. Whether reality, with its myriad of cues which are not directly relevant to a particular learning objective, is a more effective aid to communication than some edited representation of reality is still debated by practitioners and researchers.

The general preference for realism in educational practice is reinforced by several theorists. Dale's (1946) "Cone of Experience" has been particularly influential. Carpenter's (1953) "Sign Similarity Hypothesis" regarding film effectiveness and Morris' (1946) "Iconicity Theory" are also supportive of the notion. Dwyer refers to these as "realism theories."

The preference is also based partly on the affective value of realistic visuals. That is, they are expected to arouse interest and gain and sustain attention. This intuitive basis is supported somewhat by findings of a general preference for complex detailed illustrations, particularly by older children and adults (French, 1952; Spaulding, 1955; Hanes, 1973).

Dember (1960) discussed basic research which also suggests that complexity is a "major determinant of attention and exploration in both animals and man" (p. 352). Yarbus (1967), Kahneman (1973), and Haber (1973) cite related evidence that complex areas of visuals attract attention. Dember (1960) also provides evidence that stimulus complexity constitutes a rewarding condition.

Travers (1964, 1970), Dwyer (1972), and others argue, alternatively, that an increase in realism will not necessarily lead to increased learning. These later researchers feel that the irrelevant cues may interfere with learning. Travers (1964) advocates that the discarding of less important information and the retention of more important should be done for the learner to assure effectiveness (pp. 380-382). Gibson agrees that line drawings render the most important visual properties of an object, "the form, shape, and proportion of its edges and surfaces." He suggests that, "If we assume that a perceiver does not need to be given all the properties of an absent object in order to know how to deal with it, but only those which are relevant or significant, it is a waste of effort to simulate them all" (1954, pp. 21-22). Hochberg also suggests that "the characteristics of a given object may be communicated better as the representational fidelity of the surrogate deteriorates" (1962, p. 30). Dwyer (1972) found evidence to support this notion under certain conditions.

Broadbent (1958, 1965) and Travers (1964, 1970) see this inability to handle excess detail as a function of the limited capacity of the human information processing system. Broadbent describes a system in which irrelevant information is "filtered" through selective attention.

Dwyer found that the type of visual illustration (abstract vs realistic) which was most effective for instruction was dependent upon the type of information to be transmitted (or, the kind of objective), whether the instruction was self-paced or externally paced, the grade level, and the level of entering behavior of the students (1972). The type of illustration most effective was also found to be dependent on the level of general ability (IQ) (1976). For certain educational objectives, Dwyer found no significant differences between verbal only and visualized treatments (1972).

The purpose of this study was to investigate the possibility that presentations which integrate realistic and abstract visualization might prove effective under more general conditions with respect to the learner, content and presentation. Two methods of integrating the visualization were investigated. The first method was with the hybrid illustration, a real color photograph into which a line drawing segment has been inserted. This provided an abstract illustration of the important relevant elements to be communicated, as well as an arousing realistic context. The second method was through the simultaneous presentation of both a line drawing and a real color photograph. In this treatment the abstract line drawing could communicate the essential information and/or facilitate looking at the realistic illustration. The realistic illustration, on the other hand, could motivate and arouse as well as communicate essential information and relate it to reality. The effectiveness of these treatments was measured for various kinds of learning objectives, different general learning ability, externally-paced and self-paced methods of instruction, and immediate and delayed retention.

Procedure

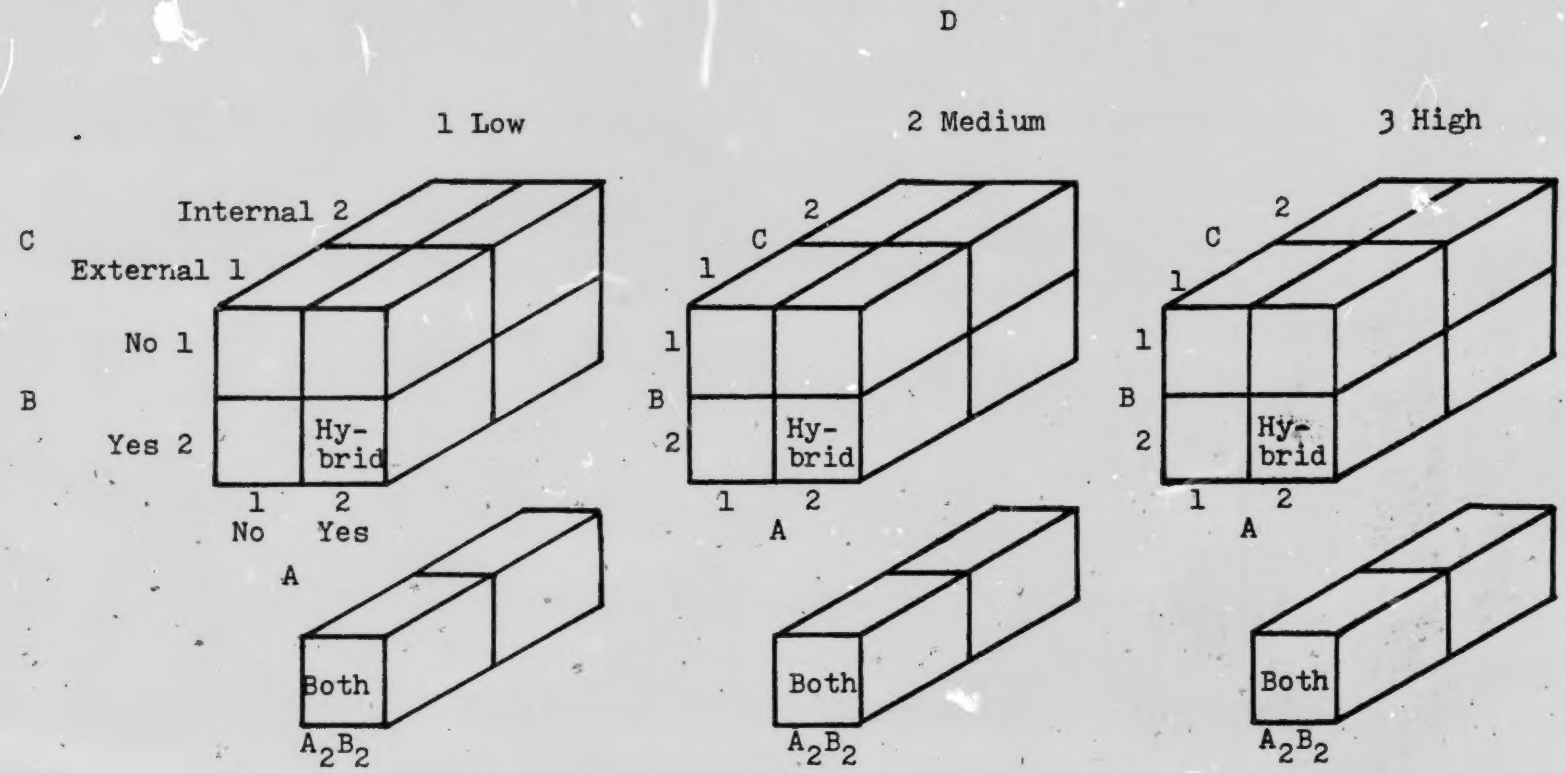
Subjects were 490 tenth grade public school students enrolled in mandatory health classes. Classes were coeducational and there was no ability grouping. The subjects were stratified into high, medium and low general ability groups on the basis of their scores from the Analysis of Learning Potential test. All subjects received a general physiology pretest prior to instruction. The instructional unit and criterion tests were those developed and used by Dwyer (1972). The 2,000 word instructional script deals with the construction and operation of the heart and is accompanied by 37 visuals, designed to illustrate concepts and relationships for which visualization is likely to be beneficial.

The criterion tests were a drawing test with 18 items and identification, terminology and comprehension tests with 20 items each. Together these four tests constituted a total criterial test. Reliabilities (KR-20) of the individual tests equal or exceed .76. The reliability of the total test is .91, according to Dwyer (1972, p. 12).

Half of the subjects received self-paced written instruction and half received externally-paced instruction by an audio tape recording (with visuals in booklet form). Within each of these two modes the students received one of the five treatments: instruction with simple line drawings, instruction with realistic color photographs (3X5 inches), instruction with both line drawings and color photographs, instruction with hybrid illustrations of realistic photograph and line drawing segments, and a control group with no visuals. Treatments were randomly assigned to subjects. All students received the four-part criterion test on the day immediately following the instruction and again after four weeks.

The two methods of integrating abstract and realistic visualization were each separately compared with the control group, the realistic visualization treatment group and the abstract visualization treatment group. These separate comparisons constituted a set of contrasts in a two-by-two matrix formed by the absence or presence of realistic visualization versus the absence or presence of abstract visualizations. This analysis model was chosen in order to facilitate identification of the anticipated additive effect of integration as an interaction of two factors.

The primary analysis involved a five-factor mixed design denoted RS_n in (A_2 @ B_2 @ C_2 @ D_3) @ J_4 (See Figure 1). Random


D
1 Low
2 Medium
3 High
C
Internal 2
External 1
No 1
B
Yes 2
Hy-brid
1 No
2 Yes
A
Both
A_2B_2
C
2
1
1
B
2
Hy-brid
1 2
A
Both
A_2B_2
C
2
1
1
B
2
Hy-brid
1 2
A
Both
A_2B_2

A = Realism

B = Abstraction

C = Mode

D = General Ability

RS_n in $(A_2 @ B_2 @ C_2 @ D_3) @ J_4$

Figure 1. Primary experimental design.

subjects were nested in cells formed by the four factors:

Factor A: Two levels of the realism variable, 1) no realism and, 2) realistic visualization.

Factor B: Two levels of the abstraction variable, 1) no abstraction and, 2) abstract visualization.

Factor C: Two levels of the presentation mode variable, 1) externally-paced and, 2) internally-paced.

Factor D: Three levels of the general ability variable, 1) Low, 2) Medium and 3) High.

All subjects are crossed with four levels of the dependent criterion variable (Factor J), 1) drawing test score, 2) identification test score, 3) terminology test score and, 4) the comprehension test score. A total criterion test score which combined the other 4 scores was analyzed separately. Scores on these tests were transformed to standard scores before analysis of results to assure comparability of the scores.

Because two methods of integrating realistic and abstract visualization were being contrasted, the entire design was analyzed twice with one oblong formed by the A and B factors (i.e., A_2 B_2) being changed A total of 30 cells are involved in the study.

A separate analysis was used to compare the two methods of integrating abstract and realistic visualization across the pacing and general ability factors. This secondary design was denoted $RS_n(A_2' @ B_2' @ C_3') @ J_4$ (See Figure 2).

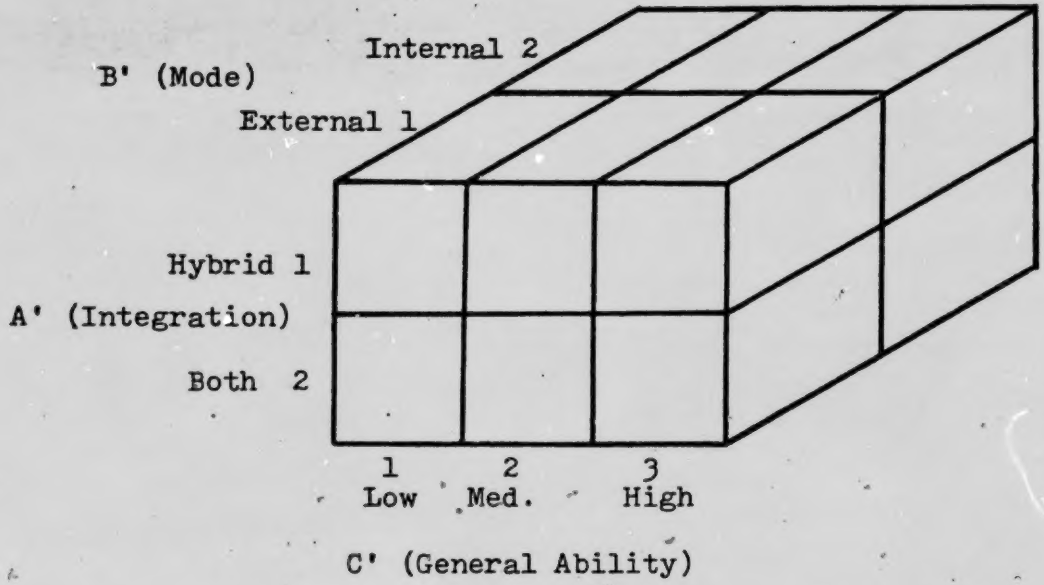

Internal 2
B' (Mode)
External 1
Hybrid 1
A' (Integration)
Both 2
1
Low
2
Med.
3
High
C' (General Ability)

Figure 2 Secondary Experimental Design

The library program ANOVR was used for the principal analysis of variance computations. This program will analyze up to four between subjects factors and up to four within subject factors in a single design. The library program FOLUP was used for the follow-up comparisons of means after significant F tests. The Tukey Wholly Significant Difference test was specified in order to control the family-wise risk of error of falsely rejecting a true null hypothesis (type I error).

Results*

In the analysis of variance procedures for both the immediate and delayed posttests, the two-way interaction effects which compared the four types of visualization were non-significant. In the analyses comparing the two methods of integrating abstract and realistic visualization the main effects of interests were similarly non-significant. Accordingly, the analysis of both the immediate and delayed posttest scores failed to reject the null hypotheses. Follow-up tests were conducted in a post-hoc investigation of other significant effects.

Immediate Posttest Findings: Hybrid Visual as the Method of Integrating Abstract and Realistic Visualization

In the analysis of variance with the four repeated measures, when the hybrid visual was the method of integrating abstract and realistic visualization, the D, BC, BJ and CJ effects produced significant F ratios.** In the analysis for the total criterion test scores only, the D and BC effects were significant. Follow-up tests were conducted on these means only.

As expected, tests on the main means of D averaged over all four tests, as well as for the total criterion test scores, indicated that high general ability students achieved significantly higher scores than medium and low ability students.* Similarly, medium general ability students achieved significantly higher scores than low ability students.

---

* Tables 1-10 in Appendix A present the standard scores for the immediate and delayed posttests.

** Significance of tests on means was at the .05 level, except where noted otherwise.

Tests on the BC means, averaged over all four tests, as well as for the total criterion test scores, indicated that when abstract visualization accompanied the instruction, students receiving externally-paced instruction achieved significantly higher scores than students receiving self-paced instruction.

Tests of the BJ means yielded no significant differences.

Tests on the CJ means indicated that students receiving self-paced instruction achieved significantly higher scores on the identification test than students who received externally-paced instruction. The results suggested these specific findings:

1. When abstract visualization accompanied instruction, externally-paced instruction was more effective than self-paced instruction, for all criterion tests and all levels of general ability.
2. Self-paced instruction was more effective than externally-paced instruction for the identification test, regardless of the type of visualization or level of general ability.

Immediate Posttest Findings: Both Types of Visuals Used Together as the Method of Integrating Abstract and Realistic Visualization

In the analysis of variance with four repeated measures, when both visuals used together was the method of integrating abstract and realistic visualization, the A, D, BC, and CJ effects produced significant F ratios. In the analysis for the total criterion test scores only, the A, D, and BC effects were significant. Follow-up tests were conducted on these means only.

---

* Significance of tests on means was at the .05 level except where noted otherwise.

Tests on the main A means, averaged over all four tests, as well as for the total criterion test scores, indicated that students who received instruction accompanied by realistic visualization achieved significantly higher scores than students who received instruction which was not accompanied by realistic visualization.

Again as expected, tests on the main D means, averaged over all four tests, as well as for the total criterion test scores, indicated that high general ability students achieved significantly higher scores than medium and low ability students. Similarly, medium general ability students achieved significantly higher scores than low ability students.

Tests on the BC means, averaged over all four tests, as well as for the total criterion test scores, yielded no significant differences.

Tests on the CJ means yielded no significant difference.

Immediate Posttest Findings: Comparison of the Two Methods of Integrating Abstract and Realistic Visualization

In the analysis with four repeated measures, when the two methods of integrating abstract and realistic visualization were compared, the C' and B'J effects yielded significant F ratios. In the analysis for the total criterion test only, the C' effect was significant.

Tests on the main C'means, averaged over all four tests, as well as for the total criterion test scores, indicated that high general ability students achieved significantly higher scores than low ability students. Similarly, medium general ability students achieved significantly higher scores than low ability students. However, high and medium general ability scores were not significantly different.

Tests on the B'J means indicated that students receiving externally-paced instruction achieved significantly higher scores on the drawing test than students who received self-paced instruction.

The results suggested these specific findings:

1. Low general ability students achieved significantly lower scores than high and medium ability students, for all criterion measures, regardless of the type of visualization or pacing. However, high and medium ability did not differ significantly in achievement.
2. Externally-paced instruction was more effective than self-paced instruction for the drawing test.

Immediate Posttest: Discussion

The findings of the immediate posttest failed to reject the null hypothesis that the various types of visualization are equally effective:

$$H_o : A_1B_1 = A_1B_2 = A_2B_1 = A_2B_2 \text{ (Hybrid)} = A_2B_2 \text{ (Both)}$$

More specifically, the findings do not support the notion that the integration of abstract and realistic visualization will generally improve the effectiveness of instruction.

However, findings from the analysis in which both types of visuals were used together to integrate abstract and realistic visualization, that instruction accompanied by realistic visualization is more effective than instruction without realistic visualization, is of related interest. This effect, in conjunction with a similar effect of the abstraction variable (B), was a necessary condition for the hypothesized additive effect of combining abstract and realistic visualization.

This finding lends general support to the "realism theories" (Dale, 1946; Morris, 1946; Carpenter, 1953) which suggest that as realism in pictures increases, so too does their instructional effectiveness. The finding may also be explained by the basic research cited by Dember (1960), Yarbus (1967), Kahneman (1973), and Haber (1973) which suggests that complex visuals serve to attract attention and increase exploration.

The same effect approached but did not reach significance in the analysis with the hybrid visual as the method of integrating abstract and realistic visualization. This failure to achieve significance may have been due to the fact that the hybrid visual contains a smaller realistic component than the complete color photo included when both types of visuals are presented together. The benefits of realistic visualization may be lost when the complete realistic visual is not used.

The finding, from the analysis with the hybrid visual as the method of integrating abstract and realistic visualization, that the presence of abstract visualization enhances externally-paced instruction, is in keeping with earlier findings by Dwyer (1972, p. 87). One possible explaning for this finding is that in externally-paced instruction students are compelled to take advantage of the edited cues in abstract visualization because of the limited time during which they may interact with them. Conversely, students receiving self-paced instruction may not be sufficiently aroused to attend to these abstract cues when time is unlimited.

The finding from the same analysis that self-paced instruction was more effective than externally-paced instruction for the kind of learning measured by the identification test is not surprising. In fact, the same general advantage is seen for all tests, although the effect was significant only for the identification test. The ability on the part of the learner to interact with the instruction at his own rate apparently enhances the kind of learning measured by the identification test.

In the analysis with only the two methods of integrating abstract and realistic visualization, externally-paced instruction resulted in greater achievement than self-paced instruction for the drawing, identification, and the terminology test. However, only the drawing test results were significant. Apparently the integration of abstract and realistic visualization enhances externally-paced instruction to a greater extent than self-paced instruction (except for the comprehension measure). This trend lends limited support to the researcher's expectancy that integration may facilitate effectiveness. One possible explanation is that students receiving externally-paced instruction were aided in extracting relevant cues in a limited time by the presence of both abstract and realistic visualization.

The performances of high, medium and low general ability students were significantly different in the two analyses with different methods of integrating abstract and realistic visualization. However, in the analysis comparing only the two methods of integration, the high and medium general ability groups' performances were not significantly different. This finding suggests that integration of realism and abstraction serves to reduce the difference in performance between high and medium general ability students. These students may be better able to take advantage of the integration of visualization because they can benefit more from the realistic component.

Delayed Posttest Findings: Hybrid Visual as The Method of Integrating Abstract and Realistic Visualization

In the analysis of variance with the four repeated measures, when the hybrid visual was the method of integrating abstract and realistic visualization, the D, BCJ, and BDJ effects produced significant F ratios. In the analysis for the total criterion test scores only, the D effect was significant. Follow-up tests were conducted on these means only.

The D means in the analysis with the four repeated measures were not of interest because they were involved in the significant BDJ interaction. However, tests of the D means for the total criterion test only indicated that high, medium, and low general ability students respectively each scored significantly higher scores than the next group.

In follow-up tests on the BCJ interaction, the BC means were tested at each level of J. These tests yielded no significant differences.

In follow-up tests of the BDJ interaction, the BD means were tested at each level of J. There were no significant differences between B_1 and B_2 at a given level of general ability (D) for any criterion tests. Significant differences in the D means at given levels of B and J indicate that the presence of abstract visualization tends to increase the difference in achievement between high and low ability students.

The results suggested these specific findings:

1. High general ability students improved moderately, though not significantly on all criterion tests when abstract visualization accompanied instruction.
2. Medium general ability students improved moderately, though not significantly, on the drawing, identification, and terminology tests when abstract visualization accompanied instruction. On the comprehension test, a trivial opposite effect occurred.
3. Low general ability students' achievement declined moderately, though not significantly, on the drawing, identification, and terminology tests when abstract visualization accompanied instruction. On the comprehension test, the opposite effect occurred.

Delayed Posttest Findings: Both Types of Visuals Used Together as the Method of Integrating Abstract and Realistic Visualization.

In the analysis of variance with four repeated measures, when both visuals used together was the method of integrating abstract and realistic visualization, the A, D, and BCJ effects produced significant F ratios. In the analysis for the total criterion test scores only, the A and D effects were significant. Follow-up tests were conducted on these means only.

Tests of the main A means, averaged over all four tests, as well as for the total criterion test scores, indicated that students who received instruction accompanied by realistic visualization achieved significantly higher scores than students who received instruction which was not accompanied by realistic visualization.

Tests on the D means, averaged over all four tests, as well as for the total criterion test scores, indicated that high general ability students achieved significantly higher scores than medium and low ability students. Similarly, medium general ability students achieved significantly higher scores than low ability students.

In follow-up tests on the BCJ interaction, the BC means were tested at each level of J. The only significant difference occurred with the comprehension test scores. These results indicated that for externally-paced instruction, students who received instruction accompanied by abstract visualization achieved significantly higher scores on the comprehension test than students who received instruction which was not accompanied by abstract visualization.

The results suggested these specific findings:

1. Students who received instruction accompanied by realistic visualization achieved significantly higher scores on all criterion measures than students who received instruction which was not accompanied by realistic visualization, regardless of the pacing or level of general ability.
2. Externally-paced instruction accompanied by abstract visualization is more effective for the kind of learning measured by the comprehension test than externally-paced instruction which is not accompanied by abstract visualization.

Delayed Posttest Findings: Comparison of the Two Methods of Integrating Abstract and Realistic Visualization

In the analysis of variance with the four repeated measures, when the two methods of integrating abstract and realistic visualization were compared, the C', B'C', and A'C'J effects yielded significant F ratios. Only the A'C'J interaction was of interest because the lower order interaction and main effect were involved in it and uninterpretable. In the analysis for the total criterion test only, the C' and B'C' effects yielded significant ratios. Accordingly, only the B'C' interaction was interpretable.

In follow-up tests on the A'C'J interaction, the A'C' means were tested at each level of J. The general trend, with one exception, was for the achievement of each ability group to improve moderately, and in one case significantly, when abstract and realistic visualization were integrated by presentation of both types of visuals simultaneously instead of by use of the hybrid visual. Low ability students achieved significantly higher scores on the drawing test when both abstract and realistic visuals were presented simultaneously. The one exception to the trend was found in the performance of the high ability students on the drawing test. The achievement of these students was higher, though not significantly, for the hybrid visual.

Tests of the A'C' means between levels of C' at each level of A' and J revealed that the presentation of both types of visuals reduced the difference in performance on the drawing test between the high and low general ability groups. No similarly strong effect was apparent for the other criterion tests.

Tests on the B'C' means for the total criterion test scores revealed no significant differences between levels of B' at a given level of general ability (C'). Similarly, there were no significant differences between the three general ability groups at B_1'. However, at B_2' (the presence of abstract visualization) high ability students scored significantly higher than medium and low ability students.

The results suggested these specific findings:

1. With one exception, all ability groups improved moderately and, in some cases significantly, when abstract and realistic visualization were integrated by presentation of both types simultaneously instead of through use of the hybrid visual.

2. The drawing test performance of low ability students was significantly improved by use of both types of visuals, instead of the hybrid visual.

3. Use of both types of visuals reduced the difference in performance on the drawing test by high, medium and low ability groups.

4. For self-paced instruction, high ability students achieved significantly higher scores than medium and low ability students.

Delayed Posttest: Discussion

As in the immediate posttest, the findings of the delayed posttest failed to reject the null hypothesis that the various types of visualization are equally effective. However, the findings do indicate that, at least for low general ability students' performance on the drawing test, presentation of both types of visuals is more effective than use of the hybrid visual as a means of integrating abstract and realistic visualization. This effect was also apparent, though not significant for other ability groups performing on other tests. Relatedly, both types presented together also reduced the difference in performance on the drawing test of high and low ability groups. These findings may possibly be explained by the increased amount of realistic visualization available when both types of visuals are presented simultaneously. Also, the findings are supportive of Hudgins (1977) who suggests possible advantages to a multi-stage use of abstract and realistic visualization.

The finding in the analysis in which both types of visuals were used together to integrate abstract and realistic visualization is more effective than instruction without realistic visualization is also of interest. This finding lends general support to the "realism theories" (Dale, 1946; Morris, 1946; Carpenter, 1953) which suggest that as realism in pictures increases, so too does their instructional effectiveness. The findings may also be explained by the basic research cited by Dember (1960), Yarbus (1967), Kahneman (1973), and Haber (1973) which suggests that complex visuals serve to attract attention and increase exploration. This effect was not present in the design in which abstract and realistic visualization were integrated by means of the hybrid visual. One possible explanation for this may be that the hybrid visual contains less realistic area than the total color photograph.

The effect of abstract visualization was inconsistent. In the analysis with the hybrid visual as the method of integrating abstract and realistic visualization, high and medium general ability students benefited moderately, though not significantly on most tests from the presence of abstract visualization. However, low ability students' scores declined moderately though not significantly, on most tests when abstract visualization accompanied instruction. Some kinds of learning by low ability students may be susceptible to long term forgetting because they were not sufficiently aroused by abstract visualization.

In the analysis in which both types of visuals were presented simultaneously to integrate abstract and realistic visualization, abstract visualization significantly enhanced the effectiveness of externally-paced instruction as measured by the comprehension test. This finding is consistent with earlier findings by Dwyer (1972). Apparently, the abstract visualization enabled students to extract relevant cues which improved learning.

Conclusions

The results of both posttests suggest that designers of visualized intruction may expect to enhance its effectiveness under general conditions by inclusion of realistic visualization. Inclusion of abstract visualization cannot be expected to improve effectiveness under general conditions and should be based on consideration of pacing, general ability and type of instructional objective.

Although the integration of abstract and realistic visualization does not appear to generally improve effectiveness, it may serve to enhance externally paced instruction and may reduce the difference in achievement between students of different general ability, particularly when both types of visualization are presented together.

REFERENCES

Broadbent, D.E. Perception and Communication. New York: Pergamon Press, 1958.

Broadbent, D.E. "Information Processing in the Nervous System." Science, 1965, 3695, 457-462.

Carpenter, C.R. "A Theoretical Orientation for Instructional Film Research." AV Communication Review, 1953, 1, 38-52.

Dale, E. Audio-visual Methods in Teaching. New York: Dryden Press, 1946.

Dember, William N. Psychology of Perception. New York: Holt, Rinehart and Winston, Inc., 1960.

Dwyer, F.M. A Guide for Improving Visualized Instruction. State College, PA: Learning Services, 1972.

Dwyer, F.M. "The Effect of IQ Level on the Instructional Effectiveness of Black-and-White and Color Illustrations." AV Communication Review, 1976, 24, 49-62.

French, J.E. "Children's Preferences for Pictures of Varied Complexity of Pictorial Pattern." The Elementary School Journal, 1952, 53, 90-95.

Gibson, J.J. "A Theory of Pictorial Perception." AV Communication Review, 1954, 2, 2-23

Haber, R.N. and Hershenson, M. The Psychology of Visual Perception. New York: Holt, Rinehart and Winston, Inc., 1973.

Hanes, M.I. "The Research on How Children Learn from Pictures." Viewpoints, 1973, 49, 11-20.

Hudgins, B.B. Learning and Thinking. Itasca, Illinois: F.E. Peacock, 1977.

Kahneman, D. Attention and Effort. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1973

Morris, C.W. Signs, Language and Behavior. New York: Prentice Hall, 1946.

Spaulding, S. "Research on Pictorial Illustrations." AV Communication Review, 1955, 3, 35-45.

Travers, R.M.W. "The Transmission of Information to Human Receivers." AV Communication Review, 1964, 12, 373-385.

Travers, R.M.W. Man's Information System. Scranton: Chandler Publishing Co., 1970.

Travers, R.M.W., ed. Research and Theory Related to Audiovisual Information Transmission. U.S. Department of Health, Education and Welfare, 1964.

Yarbus, A.L. Eye Movements and Vision. New York: Plenum Press, 1967.

TABLE 1

IMMEDIATE POSTTEST STANDARD SCORES, DRAWING TEST

		Externally-Paced General Ability Groups			Self-Paced General Ability Groups		
		Low	Medium	High	Low	Medium	High
A_1B_1	$\bar{X}$	42.22	49.00	51.33	44.11	45.56	47.89
(Control)	s	4.52	10.61	10.19	6.74	8.53	7.08
A_1B_2	$\bar{X}$	48.11	55.22	57.89	43.67	48.11	48.78
(Line Drawing)	s	7.77	8.86	2.76	6.04	8.15	8.50
A_2B_1	$\bar{X}$	48.67	48.11	51.33	43.56	51.78	58.78
(Real Photograph)	s	10.98	8.04	10.92	9.89	8.53	11.52
A_2B_2	$\bar{X}$	43.67	59.56	58.89	41.67	49.44	58.22
(Hybrid)	s	5.55	12.00	10.98	3.54	8.95	10.02
A_2B_2	$\bar{X}$	52.22	56.33	53.56	47.44	51.78	53.56
(Both)	s	11.41	7.86	12.88	8.57	7.82	6.89

n=9

TABLE 2

IMMEDIATE POSTTEST STANDARD SCORES, IDENTIFICATION TEST

		Externally-Paced General Ability Groups			Self-Paced General Ability Groups		
		Low	Medium	High	Low	Medium	High
A_1B_1	$\overline{X}$	42.89	48.22	51.89	48.78	43.56	52.67
(Control)	s	5.3	9.6	13.2	5.4	6.4	10.3
A_1B_2	$\overline{X}$	48.56	54.56	58.00	43.89	47.00	47.22
(Line Drawing)	s	5.0	10.0	10.7	6.0	10.9	10.7
A_2B_1	$\overline{X}$	49.56	50.11	54.56	43.22	50.78	57.56
(Real Photograph)	s	10.0	11.4	8.3	7.8	8.5	5.2
A_2B_2	$\overline{X}$	46.56	56.44	62.11	43.22	45.22	60.33
(Hybrid)	s	7.5	11.6	10.5	3.0	9.0	12.1
A_2B_2	$\overline{X}$	49.89	57.11	49.78	45.33	53.22	55.44
(Both)	s	8.6	8.7	12.3	9.4	10.7	8.6

n=9

TABLE 3

IMMEDIATE POSTTEST STANDARD SCORES, TERMINOLOGY TEST

		Externally-Paced General Ability Groups			Self-Paced General Ability Groups		
		Low	Medium	High	Low	Medium	High
A_1B_1	$\overline{X}$	45.33	51.67	48.22	48.56	49.11	56.78
(Control)	s	6.4	7.3	11.3	8.7	9.9	8.6
A_1B_2	$\overline{X}$	48.56	50.44	59.22	43.78	47.00	49.11
(Line Drawing)	s	6.4	7.6	14.8	6.0	11.9	7.8
A_2B_1	$\overline{X}$	44.00	52.89	51.00	47.22	50.78	60.67
(Real Photograph)	s	7.4	9.7	12.4	5.3	10.0	7.0
A_2B_2	$\overline{X}$	44.56	55.44	60.11	41.78	50.11	58.67
(Hybrid)	s	7.8	10.5	16.1	5.04	10.9	10.8
A_2B_2	$\overline{X}$	46.33	52.78	52.89	44.11	52.44	58.56
(Both	s	5.8	10.7	10.4	4.7	8.3	12.1

n=9

TABLE 4

IMMEDIATE POSTTEST STANDARD SCORES, COMPREHENSION TEST

		Externally-Paced General Ability Groups			Self-Paced General Ability Groups		
		Low	Medium	High	Low	Medium	High
A_1B_1	$\overline{X}$	44.44	51.67	50.78	47.22	51.56	53.78
(Control)	s	4.7	8.8	10.8	8.5	6.3	12.0
A_1B_2	$\overline{X}$	50.22	50.22	52.33	46.78	47.89	50.00
(Line Drawing)	s	6.7	10.9	14.2	6.6	8.0	5.3
A_2B_1	$\overline{X}$	45.78	49.22	56.22	50.22	49.44	58.66
(Real Photograph)	s	9.1	9.7	9.9	9.2	8.4	15.3
A_2B_2	$\overline{X}$	44.33	50.78	54.56	42.67	49.78	57.78
(Hybrid)	s	4.7	10.6	16.3	5.4	8.4	12.4
A_2B_2	$\overline{X}$	45.89	51.00	53.44	46.44	51.44	59.89
(Both)	s	8.4	7.4	8.6	6.1	12.3	9.8

n=9

TABLE 5

IMMEDIATE POSTTEST STANDARD SCORES, TOTAL CRITERION TEST

		Externally-Paced General Ability Groups			Self-Paced General Ability Groups		
		Low	Medium	High	Low	Medium	High
A_1B_1	$\bar{X}$	42.22	50.11	51.00	46.44	46.44	52.89
(Control)	s	4.3	7.9	11.8	6.4	4.3	9.4
A_1B_2	$\bar{X}$	48.78	53.44	58.56	43.22	47.00	48.44
(Line Drawing)	s	5.3	8.7	11.7	5.9	10.0	8.6
A_2B_1	$\bar{X}$	46.78	49.89	53.89	44.89	50.78	60.78
(Real Photograph)	s	10.3	9.2	10.9	7.3	9.2	9.1
A_2B_2	$\bar{X}$	43.78	57.22	61.00	40.67	48.33	60.67
(Hybrid)	s	5.3	11.1	14.8	2.3	9.3	12.2
A_2B_2	$\bar{X}$	48.78	55.56	53.11	45.11	52.67	57.89
(Both)	s	6.5	9.0	12.0	5.4	9.1	5.3

n=9

TABLE 6

DELAYED POSTTEST STANDARD SCORES, DRAWING TEST

		Externally-Paced General Ability Groups			Self-Paced General Ability Groups		
		Low	Medium	High	Low	Medium	High
A_1B_1	$\bar{X}$	42.89	51.11	50.22	46.11	45.78	54.22
(Control)	s	3.59	10.90	10.35	7.15	6.53	7.69
A_1B_2	$\bar{X}$	49.89	51.89	57.44	41.56	49.11	53.44
(Line Drawing)	s	9.74	7.77	9.80	2.88	10.41	11.35
A_2B_1	$\bar{X}$	50.11	49.89	51.67	45.89	52.11	57.11
(Real Photograph)	s	8.58	7.29	9.38	9.55	9.49	13.06
A_2B_2	$\bar{X}$	40.89	53.22	58.89	41.22	50.78	59.44
(Hybrid)	s	2.15	9.85	13.47	2.86	10.35	13.05
A_2B_2	$\bar{X}$	51.22	53.33	53.78	47.11	51.67	55.22
(Both)	s	10.78	8.17	12.36	6.58	7.40	14.21

n=9

TABLE 7

DELAYED POSTTEST STANDARD SCORES, IDENTIFICATION TEST

•		Externally-Paced General Ability Groups			Self-Paced General Ability Groups		
		Low	Medium	High	Low	Medium	High
A_1B_1	$\overline{X}$	46.00	49.78	48.67	48.11	45.56	52.22
(Control)	s	5.92	16.84	13.76	8.80	5.50	8.18
A_1B_2	$\overline{X}$	44.33	50.67	58.67	47.33	45.56	56.89
(Line Drawing)	s	3.81	7.52	12.84	5.81	6.46	8.59
A_2B_1	$\overline{X}$	52.78	49.44	53.78	45.56	50.33	55.11
(Real Photograph)	s	12.71	5.27	12.01	8.71	10.61	13.49
A_2B_2	$\overline{X}$	45.33	51.11	55.22	45.89	51.00	58.78
(Hybrid)	s	3.28	8.91	14.11	8.55	10.05	11.33
A_2B_2	$\overline{X}$	48.22	52.33	53.00	45.33	51.78	61.67
(Both)	s	5.93	8.63	12.01	6.86	7.89	13.45

n=9

TABLE 8

DELAYED POSTTEST STANDARD SCORES, TERMINOLOGY TEST

		Externally-Paced General Ability Groups			Self-Paced General Ability Groups		
		Low	Medium	High	Low	Medium	High
A_1B_1	$\overline{X}$	44.56	53.33	50.44	44.11	47.00	50.56
(Control)	s	5.88	11.47	7.09	8.36	8.96	12.51
A_1B_2	$\overline{X}$	44.11	45.67	55.89	43.89	51.67	54.67
(Line Drawing)	s	3.22	6.76	16.79	5.58	7.87	6.82
A_2B_1	$\overline{X}$	46.00	48.33	58.56	48.11	52.78	58.00
(Real Photograph)	s	5.29	9.19	16.79	9.89	9.56	10.81
A_2B_2	$\overline{X}$	44.78	50.89	49.00	45.89	45.67	61.56
(Hybrid)	s	6.22	8.24	8.05	4.43	6.95	12.61
A_2B_2	$\overline{X}$	46.56	51.00	52.89	46.33	47.44	67.22
(Both)	s	8.32	9.03	6.85	7.04	7.88	11.27

n=9

TABLE 9

DELAYED POSTTEST STANDARD SCORES, COMPREHENSION TEST

		Externally-Paced General Ability Groups			Self-Paced General Ability Groups		
		Low	Medium	High	Low	Medium	High
A_1B_1	$\bar{X}$	41.78	45.67	50.78	44.33	49.89	51.78
(Control)	s	5.04	5.68	8.57	6.20	11.19	9.19
A_1B_2	$\bar{X}$	51.56	49.56	56.78	50.00	47.33	50.56
(Line Drawing)	s	10.08	4.93	17.69	4.64	16.19	8.97
A_2B_1 (Real	$\bar{X}$	45.67	46.89	51.56	45.44	53.67	55.56
Photograph)	s	8.34	7.87	10.97	9.22	10.46	16.00
A_2B_2	$\bar{X}$	47.00	48.89	52.33	46.44	47.67	57.67
(Hybrid)	s	7.40	10.08	7.19	7.35	10.19	10.36
A_2B_2	$\bar{X}$	51.56	51.56	51.77	46.67	50.67	66.33
(Both)	s	7.94	13.54	8.50	10.04	11.64	12.48

n=9

TABLE 10

DELAYED POSTTEST STANDARD SCORES, TOTAL CRITERION TEST

		Externally-Paced General Ability Groups			Self-Paced General Ability Groups		
		Low	Medium	High	Low	Medium	High
A_1B_1	$\bar{X}$	42.44	50.11	50.00	45.11	46.00	53.00
(Control)	s	3.50	12.40	10.25	7.06	6.87	8.03
A_1B_2	$\bar{X}$	47.00	49.67	58.89	44.44	48.00	55.11
(Line Drawing)	s	6.16	5.77	15.23	4.56	10.72	9.75
A_2B_1	$\bar{X}$	49.00	48.44	54.44	45.33	52.56	57.78
(Real Photograph)	s	9.70	5.27	12.88	9.49	10.27	13.20
A_2B_2	$\bar{X}$	42.89	51.44	55.56	43.33	48.89	61.33
(Hybrid)	s	1.96	2.92	12.65	4.39	7.75	12.60
A_2B_2	$\bar{X}$	49.22	52.89	53.44	45.33	50.78	64.44
(Both)	s	7.61	8.85	9.30	6.04	9.02	14.15

n=9

Efficiency of Students' Achievement Using Black/White and Color Coded Learning and Test Materials

Richard J. Lamberski
Director-University Media Services
Assistant Professor -Educational Media and Technology
Boston University

and

Dennis M. Roberts
Associate Professor- Educational Psychology
The Pennsylvania State University

Research and Theory Division
Association for Educational Communications & Technology
New Orleans, Louisiana
Wednesday- March 7,1979

Efficiency of Students' Achievement Using Black/White and Color Coded Learning and Test Materials

Abstract

The purpose of this study was to compute and analyze descriptive indices of efficiency assessing the relative effectiveness of a verbal and visual color code applied to self-paced learning and testing materials. The sample consisted of 176 college subjects randomly assigned to one of four treatment conditions receiving black/white (B) or color coded (C) materials: B learning and testing materials, B learning materials but C testing materials, C learning and testing materials, or C learning materials but B testing materials.

Materials consisted of a 21 page learning booklet on the human heart and an 80 item test measuring different levels of cognitive knowledge; content was identical except that in some treatment conditions a color code having meaningful physical form and associative value was applied. Subjects read through the learning booklet followed immediately by a test; six weeks later the identical test was administered to assess delayed retention. Time on learning and time on tests was recorded.

Calculation of three related efficiency formulas (achievement per unit of time) provided the descriptive data for discussion. Data indicated that while color coded learning materials facilitated achievement, additional time for acquisition and for retrieval was necessary. This primary relationship, in addition to secondary relationships was further discussed in light of each efficiency derivation.

Efficiency of Students' Achievement Using Black/White and Color Coded Learning and Test Materials

Purpose and Rationale

Research during the instructional media era of the 1950's and 1960's was comparative in nature (Torkelson, 1978) and sought to identify *device* usefulness rather than to identify the parameters impacting on elements of the *message* (Dwyer, 1972; 1978). Questions posed were to justify using media rather than to optimize learning. It was this later need to investigate those message design elements which in combination or limitation may facilitate optimal (effective and efficient) learning that this study sought to supply additional evidence for. Specifically this investigation sought to clarify the cognitive value of color when used as a coding device in the design of self-paced learning and testing materials.

In surveying the color literature researchers (Lamberski, 1972; 1979; Berry, 1974; Chute, 1978) have concluded that the significance of color as a design variable has typically eluded researchers who have attempted to define its *unique* role rather than its possible *interrelated* role in the learning process. In many studies the use of color was often superseded by other variables or not properly isolated by experimental design to assess its effectiveness.

Exploratory investigations in cognitive learning have found that color may facilitate or inhibit learning when used as an instructional cue or code in message design (Kauffman and Dwyer, 1974; Lamberski, 1975; Berry, 1975; 1976). While there is support for the affective preference for color and for the physiological and developmental influence of color, it is evident that the cognitive value of color to the learning process must be clarified further (Dwyer, 1978; Lamberski, 1979). Thus, the purpose of this study was to compute and analyze descriptive indices of efficiency assessing the relative effectiveness of a verbal and visual color code applied to self-paced learning and testing materials. It was hypothesized that when both learning and testing materials utilized a color code, learning would be most efficient (greatest performance or achievement per unit of time).

Method

Sample

The sample consisted of 176 (39 male and 137 female) student volunteers from an instructional media course at The Pennsylvania State University.

Instructional Materials

The learning material (Dwyer and Lamberski, 1977) was a revision of a 2,000 word prose text, with accompanying simple line drawing visualization and verbal labels, originally published by Dwyer (1972). The content of the material consisted of the parts and internal processes of the human heart. Two versions of the 21 page learning booklet were developed; one was entirely black/white (B) while the other one had a color code (C) applied.

The color code was applied only to relevant (central) concepts or functions being presented on a particular page with no color coding when the concept or function was redundant (peripheral) to other instruction being presented. The color code application in the booklet was for both the verbal instruction (the ink color of the printed concept or function word) and accompanying visual illustration (the identical ink color of the printed concept or function label, arrow, and shaded area). The six colors and black formed categories which related similar concepts and functions or discriminated between the other concept or function words. Thus, the color code had both a physical form (visual with cues) and an associative value (verbal with meaning). Such a color code application has been applied in various primary reading investigations (Hinds and Dodds, 1968), and more recently color visualization has been applied and found to have an associative value (Winn, 1976; 1977).

The test material (Dwyer and Lamberski, 1977) was a revision of the battery of tests originally published by Dwyer (1972). The materials consist of four 20-item subtests (summed together equal 80 items) assessing hierarchical cognitive knowledge of drawing, identification, terminology, and comprehension. Two forms of the test were prepared, one to follow each of the black/white (B) and color coded (C) learning material formats.

All materials were piloted prior to the present investigation. Test instrument reliability coefficients for data in this study, determined by the Kuder-Richardson formula 20 as being between .74-.88 for the four respective subtests (Lamberski, 1979), compared quite favorably with coefficients found in investigations using Dwyer's original materials (Dwyer, 1972).

Procedure

Participants were randomly assigned to one of four treatment conditions: each treatment represented a crossing of the two levels - black/white (B) or color coded (C) - for each of the independent variables - learning and testing materials. The resulting cells in the design therefore were: B learning and B test materials, B learning but C test materials, C learning and C test materials, and C learning but B test materials. Subjects read the learning booklet immediately after which they were administered a test; six weeks later the identical test was administered to assess delayed retention. Time on the learning task and time on both retention tests were recorded by each subject. There were N = 44 subjects in each treatment condition.

Results and Discussion

Three descriptive indices of instructional efficiency were computed whereby the number of correct items on the immediate or delayed tests were divided by a corresponding time measure. Thus, the dependent variables for subjects in each treatment condition were: the mean achievement score (Table 1) which represents the number of correct items (80 items total) on each respective retention test; the mean learning time (Table 2) which represents how long it took to complete the learning materials (expressed in minutes); and the mean test time (Table 3) which represents how long it took to complete the testing materials (expressed in minutes) on each respective retention test. Thus, the derived mean efficiency values for the three descriptive indices (Tables 4,5, & 6) represent mean number correct per minute on each respective retention test. Data are visually summarized in Figures 1, 2, and 3.

Calculation of the three related efficiency formulas (achievement per unit of time) provided evidence which indicated that while color coded learning materials facilitated student achievement, additional time for acquisition and for retrieval was necessary. Data showed higher efficiencies for immediate test results; this simply reflects that recall was poorer at delayed retention testing. Also, performance or achievement was more efficient when black/white tests were used. Data did not support the hypothesis that the most efficient instructional strategy was to pair color coded learning materials with color coded tests. While the second efficiency index (Figure 2, Table 5) using the test time measure was computed primarily for interest, the first (Figure 1, Table 4) and the third (Figure 3, Table 6) efficiency index values should be viewed as more realistic since learning time or the combination of learning and testing time will normally far outweigh testing time alone.

The findings might mean that color coding facilitated the quality and the quantity of the encoding and the decoding processes but because the discrimination amongst numerous colors was a more time consuming process, the increased encoding and decoding time lowers the efficiency index. However, the data did suggest that color coding of learning materials could be an efficient instructional strategy when coupled with the simpler black/white test. This might be an artifact of the current data though, since testing time was similar to learning time. Certainly, a better situation would be to extend learning time considerably (perhaps 10 fold or more) while holding testing time constant. That should provide a more realistic test of the original efficiency hypothesis.

* * * * * * * *

References

Berry, L.H. An exploratory study of the relative effectiveness of realistic and non-realistic color in visual instructional materials. Doctoral Dissertation, The Pennsylvania State University, 1974.

Berry, L.H. An investigation of the effectiveness of realistic and non-realistic color in visualized instruction. A paper presented at the AECT Convention, Dallas, TX, 1975.

Berry, L.H. Interactive effects of color realism and learner I.Q. on visualized instruction. A paper presented at the AECT Convention, Anaheim, CA, 1976.

Chute, A.G. An experimental study investigating the comparative effectiveness of a color and a monochrome version of an instructional film presentation on incidental and task-relevant discrimination learning. Doctoral Dissertation, University of Wisconsin at Madison, 1978.

Dwyer, F.M., Jr. A Guide for Improving Visualized Instruction. State College, PA: Learning Services, 1972.

Dwyer, F.M., Jr. Strategies for Improving Visual Learning. State College, PA: Learning Services, 1978.

Dwyer, F.M., Jr. and Lamberski, R.J. The Human Heart: Instructional Unit and Test Materials. 1977. Supplementary volume to Lamberski (1979).

Hinds, L.R. and Dodds, W. Words in color: two experimental studies. Journal of Typographic Research, 1968, 2(1), 43-56.

Kauffman, S.P. and Dwyer, F.M., Jr. Effectiveness of cartoons and photographs in in-service training. California Journal of Educational Research, 1974, 25(4), 197-204.

Lamberski, R.J. An exploratory investigation on the instructional effect of color and black and white cueing on immediate and delayed retention. Master's Thesis, The Pennsylvania State University, 1972.

Lamberski, R.J. An exploratory study in maximizing retention by utilizing black/white and color coding in visualized instruction. P paper presented at the AECT Convention, Dallas, TX, 1975.

Lamberski, R.J. The effect of a verbal and visual color code on self-paced instruction and testing for retention on different tasks. Doctoral Dissertation, The Pennsylvania State University, 1979.

Torkelson, G.M. One quarter century: evolution of theory and research. AV Communication Review, 1978, 24(4), 317-358.

Winn, W.D. The structure of multiple free associations towards black and white pictures and color pictures. AV Communication Review, 1976, 24, 273-293.

Winn, W.D. The effect of the removal of common meaning on the structure of free associations to black and white and color pictures. A paper presented at the AECT Convention, Miami Beach, FL, 1977.

* * * * * * * * *

Contact Addresses

Richard J. Lamberski
Boston Univ. Media Services
765 Commonwealth Ave, Rm 1454
Boston, MA 02215

Dennis M. Roberts
418 Carpenter Building
The Pennsylvania State Univ.
University Park, PA 16802

TABLE 1 - Mean Achievement Scores

Treatment

Retention	B/B	B/C	C/C	C/B	
Immediate	52.64	51.34	61.38	58.77	56.03
Delayed	39.91	39.28	44.20	43.48	41.72
	46.28	45.31	52.79	51.13	48.88

TABLE 2 - Mean Learning Time

Treatment

B/B	B/C	C/C	C/B	
14.11	14.93	20.18	16.50	16.43

TABLE 3 - Mean Test Time

Treatment

Retention	B/B	B/C	C/C	C/B	
Immediate	25.47	26.12	29.12	25.89	26.65
Delayed	22.89	23.08	28.25	24.24	24.62
	24.18	24.60	28.69	25.07	25.64

Table Notes:

B/B = Black/white learning and test materials
B/C = Black/white learning materials but color test materials
C/C = Color learning materials and test materials
C/B = Color learning materials but black/white test materials

Tables 2 and 3: Values represent minutes.

Figure 1 Efficiency: Achievement/Learning Time

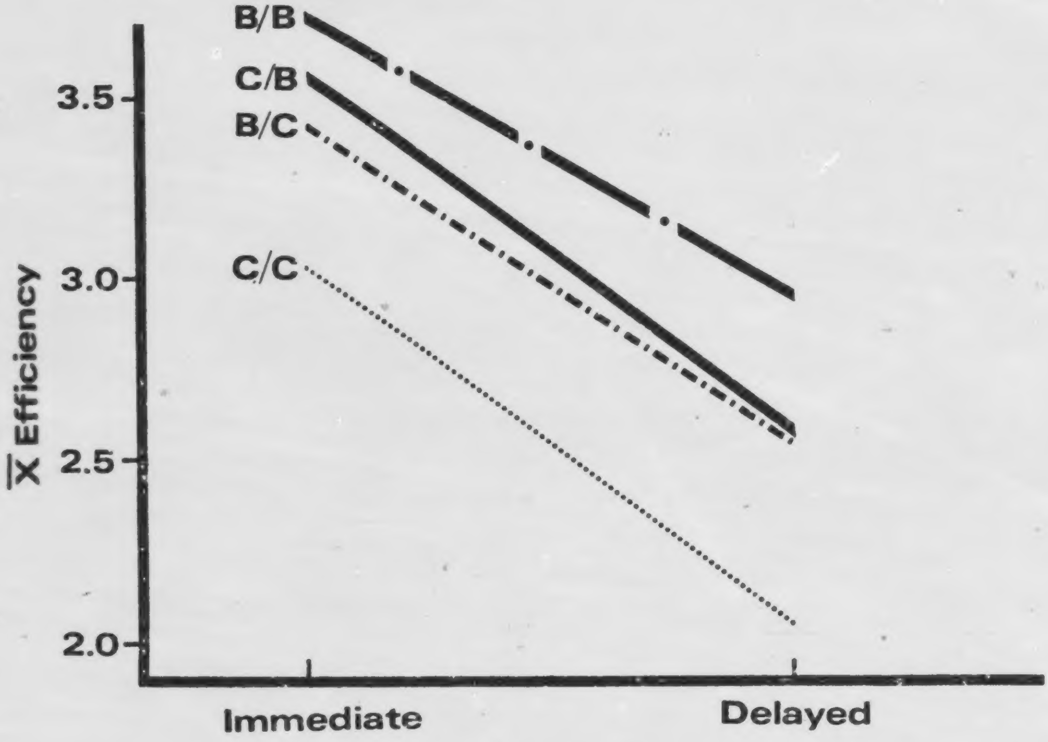


TABLE 4 - Efficiency: Achievement/Learning Time

Retention	Treatment B/B	Treatment B/C	Treatment C/C	Treatment C/B	
Immediate	3.731	3.439	3.042	3.562	3.443
Delayed	2.828	2.631	2.190	2.635	2.571
	3.280	3.035	2.616	3.099	3.007

Note: Values represent mean number correct per minute.

Figure 2 Efficiency: Achievement/Test Time

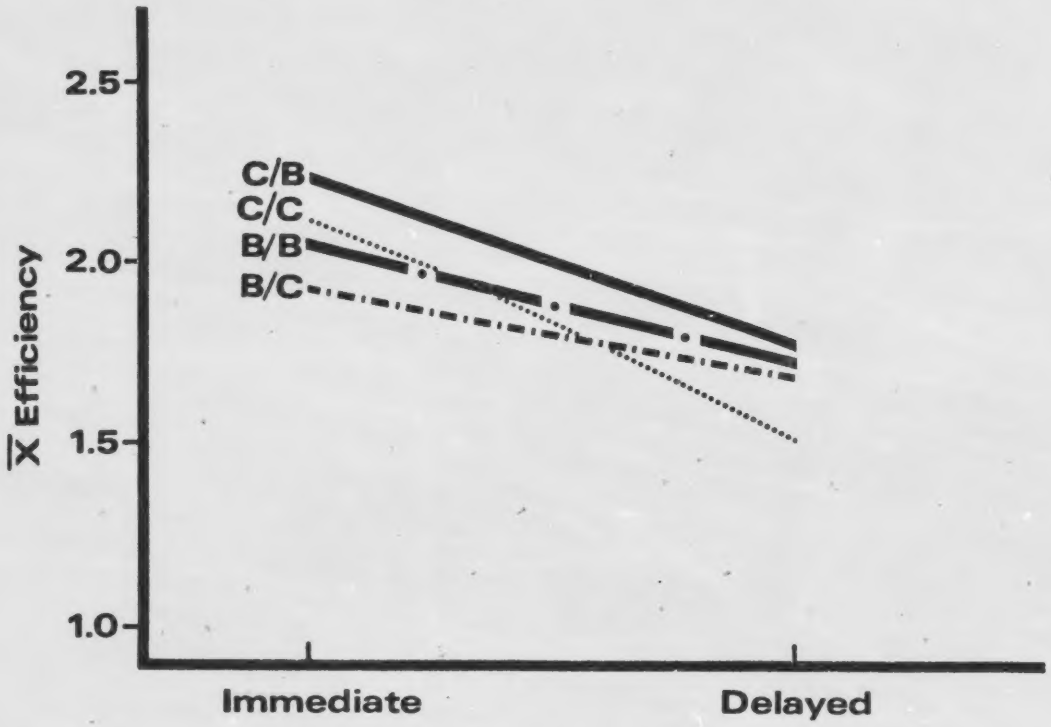


TABLE 5 - Efficiency: Achievement/Test Time

Retention	B/B	B/C	C/C	C/B	
	Treatment				
Immediate	2.067	1.966	2.108	2.270	2.103
Delayed	1.744	1.702	1.565	1.794	1.701
	1.906	1.834	1.837	2.032	1.902

Note: Values represent mean number correct per minute.

Figure 3 Efficiency: Achievement/Learning & Test Time

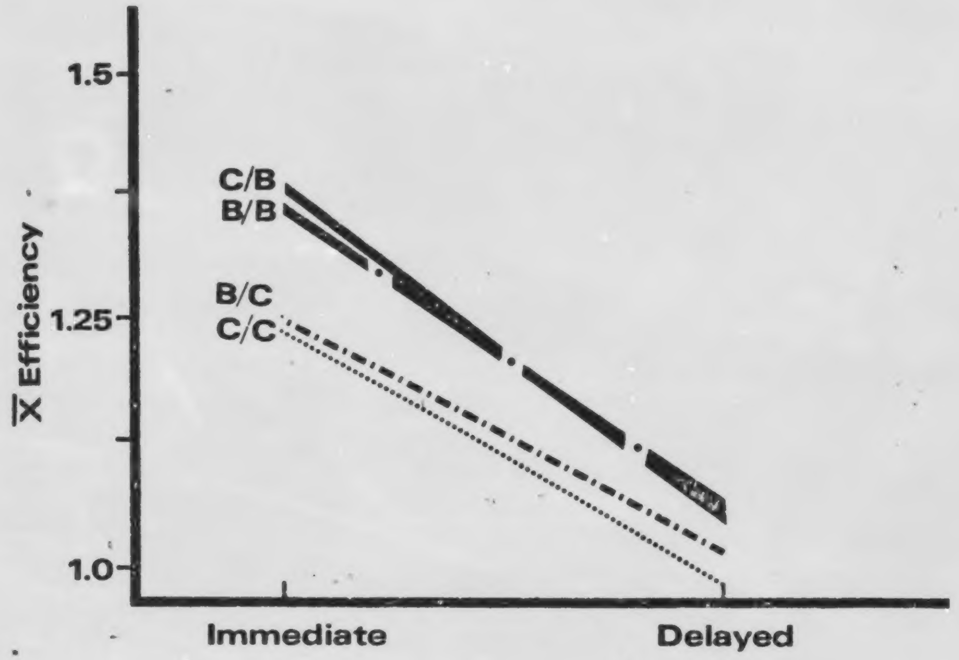


TABLE 6 - Efficiency: Achievement/Learning & Test Time

Retention	B/B	B/C	C/C	C/B	
			Treatment		
Immediate	1.330	1.251	1.245	1.386	1.303
Delayed	1.079	1.033	0.913	1.067	1.023
	1.205	1.142	1.079	1.227	1.163

Note: Values represent mean number correct per minute.

IR007340

THE EFFECT OF MUSIC AND SOUND EFFECTS ON THE LISTENING COMPREHENSION OF FOURTH GRADE STUDENTS

Raymond E. Mann, Ed.D.
Director, Education Media Center
College of Education
NORTH TEXAS STATE UNIVERSITY

March, 1979

Mann, Raymond E., The Effect of Music and Sound Effects on the Listening Comprehension of Fourth Grade Students. North Texas State University, March, 1979, 27pp., 7 tables, bibliography.

The purpose of this study was to determine if the addition of music and sound effects to recorded stories increased the comprehension and retention of information presented on tape for fourth grade students. Two versions of four narrated stories were recorded--one version with music and sound effects, the other included the narration only. A listening comprehension test was administered after listening to each of the stories and followed by a retention test two weeks later. Students listening to the music and sound effects versions scored significantly higher on both the listening comprehension and retention tests.

INTRODUCTION

Listening is one of the first activities in the life of every individual. Research has indicated that newborn infants can make "fine discriminations in sounds. Sounds have meaning for babies, even before they have learned language" (16).

Children learn to speak primarily through listening to others. However, primary children have difficulty in listening effectively to dialogues encountered in storytelling situations or in following instructions (3, 4, 14, 17,). Many of the reasons why children have difficulty in listening have not been identified.

The importance of listening as a distinct learning process has been stressed by psychologists and educators for some time. Duker states that listening "has always been and continues to be the most widely used human means of receiving information" (7) from others. An individual's listening activities are basic in order to learn to "speak, read, and write" (6).

Studies by Rankin and Wilt indicate children spend over fifty per cent of their school day listening (9, 19) and students who score high on listening comprehension tests normally do better in school (15). Studies have also shown that students can be taught to listen more effectively (11).

Even though listening has received much recognition and the importance of listening has been stressed by many educators, very little research has been done on what factors contribute to one's ability and willingness to listen. Although numerous articles about listening have been written, few are based on systematic research on listening as an information gathering psychological process.

Also important is the lack of research on the numerous instructional presentation modes, all of which have an effect on the listening process.

> In view of the increasing instructional use of a wide range of teaching materials and presentation media--such as video cassettes, audiovisual teaching kits, audio cassettes, etc.--the necessity for such systematic investigation is readily apparent particularly in that some media may be more or less effective than others (20, p. 32).

The increase of media utilization in American schools during the past twenty-five years has been unprecedented and the potential educational significance of these materials is tremendous. However, the instructional potential of media can never be achieved without a solid foundation of research and understanding of children's listening competence (20).

Statement of the Problem

The problem of this study was to determine the effect, if any, of music and other sound effects on listening comprehension and recall of fourth grade students.

Purposes of the Study

The purposes of this study were:

1. To determine if fourth grade students' comprehension of stories recorded with music and sound effects was greater than straight narrative stories.

2. To determine if fourth grade students' recall of information from stories recorded with music and sound effects was greater than straight narrative stories.

3. To determine if fourth grade students' reading ability was a factor in comprehension and recall of information from stories recorded with music and sound effects or those stories recorded without music and sound effects.

Hypotheses

The following hypotheses were tested by statistical analysis of the data collected:

1. There will be no significant difference in the comprehension of information by fourth grade students who listened to narrated stories with music and other sound effects and by those who listened to the same stories without music and other sound effects.

2. There will be no significant difference in the comprehension of information by fourth grade students with low reading abilities who listened to narrated stories with music and other sound effects and by those who listened to the stories without music and other sound effects.

3. There will be no significant difference in the comprehension of information by fourth grade students with average of medium reading abilities who listened to narrated stories with music and other sound effects and by those who listened to the stories without music and other sound effects.

4. There will be no significant difference in the comprehension of information by fourth grade students with high reading abilities who listened to narrated stories with music and other sound effects and by those who listened to the stories without music and other sound effects.

5. There will be no significant difference in the retention of information by fourth grade students who listened to narrated stories with music and other sound effects and by those who listened to the stories without music and other sound effects.

6. There will be no significant difference in the retention of information by fourth grade students with low reading abilities who listened to narrated stories with

music and other sound effects and by those who listened to the stories without music and other sound effects.

7. There will be no significant difference in the retention of information by fourth grade students with average reading abilities who listened to narrated stories with music and other sound effects and by those who listened to the stories without music and other sound effects.

8. There will be no significant difference in the retention of information by fourth grade students with high reading abilities who listened to narrated stories with music and other sound effects and by those who listened to the stories without music and other sound effects.

METHODOLOGY

Population

The subjects involved in this study were 107 fourth grade students enrolled in two elementary schools in a public school district in north-central Texas. The students represented all fourth grade students enrolled in both elementary schools. The school district encomposses a small city and the surrounding rural area. The two schools used in the study were selected because they each had self-contained classrooms where all activities could be conducted.

Procedures for Conducting the Study

To eliminate the possibility that differences in the composition of treatment groups might bias the study, an experimental design was developed for use with this study which allowed the entire sample to function as the experimental and the control group. The five classes of fourth grade students available for the study were arbitrarily assigned to Group A or Group B with an attempt to have approximately equal numbers in each group. Each of the five classes listened to four taped stories ranging from five to ten minutes in length. Each group received two of the stories presented in the control format ant two stories in the experimental format. The net result assured the the experimental and control groups were, in fact, equal. Group A and Group B were never compared against one another; only the treatment scores were compared. The two audio presentation formats utilized in the study were:

1. Four stories were narrated and recorded on monaral cassette for playback.

2. The four stories were revised by including music and sound effects in a radio drama format though the narration remained unchanged.

The study was conducted during the month of May, 1978, at the same time each day. The students were informed at the beginning of the first day that they would be listening to four stories, one each day for four days, and would be asked to answer a twenty item multiple-choice test over each of the stories.

Two weeks after listening to the fourth story, all students were administered a recall test to measure retention of information from the stories. The recall test was composed of ten multiple-choice questions selected from each of the four comprehension tests. Thus the recall test was a forty item test. No discussion of the stories was conducted between the time of the first tests and the final recall test. Only the scores of those students who were present on all four days and took the recall test were included in the final analyses.

The week preceding the administration of the listening comprehension tests, all subjects in the study had been administered the California Test of Basic Skills (CTBS) along with all students in the school system. The Reading Vocabulary and Reading Comprehension scores were combined in the analyzed data to give a total "Reading" score expressed as a "National Percentile" score. The "National Percentile" score was utilized to divide the total subject population into three groups of approximately equal size representing the upper, middle, and lower thirds of the test population.

Collection of the Data

Each subject had test scores from each of the four comprehension tests and the four part recall test. The tape treatments that each student received were recorded and utilized to separate scores into "music and sound effects" and "narration only" columns. The net result was a total of two scores for each subject for each of the two treatments. There were thirty-six subjects in the high reading category, thirty-seven subjects in the medium category, and thirty-four subjects in the low category.

Treatment of the Data

The data for this study were obtained from the four twenty item multiple-choice tests for comprehension, the forty item multiple-choice test for recall, and the CTBS National Reading Percentile. The data from the comprehension tests and the recall test were analyzed separately utilizing the same procedures. The data from the comprehension test scores were analyzed by a two factor analysis of variance with repeated measures. When an overall significant difference was found between groups, a multiple comparison test was run to determine which pairs of treatment levels were significantly different. Within each reading level group, an analysis of variance test was conducted to determine if a significant difference in treatments was found. The same procedures were used in the analysis of the data from the recall test. The level of significance for this study was the .05 level.

The statistical data was analyzed by the Computer Center at North Texas State University and is presented in table form. From the findings, conclusions were drawn, educational implications stated, and recommendations made.

PRESENTATION OF THE FINDINGS

The cell data for the comprehension test scores are recorded in Table I. The population is divided into high, medium, and low reading ability, and includes scores for the total population. In addition to the number of subjects (N) in each cell, the mean (M) and standard deviation (SD) for each treatment combination, and the combined total treatment scores are included.

TABLE I

NUMBER OF SUBJECTS, MEANS, AND STANDARD DEVIATIONS FOR THE COMBINED SCORES OF COMPREHENSION TESTS AND THE COMBINED MEANS OF THE MAIN VARIABLES

Reading Group		Music and Sound Effects	Narration Only	Group Means
High	N M SD	36 35.1389 2.9096	36 34.6667 3.3722	34.9028
Medium	N M SD	37 31.6486 4.4921	37 29.8649 5.2766	30.7568
Low	N M SD	34 24.1471 7.6441	34 21.7647 8.0756	22.9559
Total	N M SD	107 30.4393 6.9678	107 28.9065 7.8398	

A comparison of the mean scores for both treatments within each group and for the total population yield a higher mean score for the music and sound effects treatment than the narration only treatment. It should also be noted that the mean scores for both treatments were higher for the high reading group and lowest for the low reading group.

TABLE II

NUMBER OF SUBJECTS, MEANS, AND STANDARD DEVIATIONS FOR THE COMBINED SCORES OF THE RECALL TEST AND THE COMBINED MEANS OF THE MAIN VARIABLES

Reading Level		Music and Sound Effects	Narration Only	Group Means
High	N	36	36	
	M	18.0278	17.0556	17.5417
	SD	1.6985	2.8380	
Medium	N	37	37	
	M	15.6486	14.3514	15.0000
	SD	3.0507	3.2678	
Low	N	34	34	
	M	10.9706	10.0000	10.4853
	SD	4.7064	4.4107	
Total	N	107	107	
	M	14.9626	13.8785	
	SD	4.4147	4.5468	

The cell data for the recall tests are given in Table II. Again, the population is divided into high, medium, and low reading ability groups, and include scores for the total population. In addition to the number of subjects (N) in each cell, the mean (M) and standard deviation (SD) for each treatment combination, and the combined total treatment scores are included.

As with the comprehension test scores, a comparison of the mean scores for both treatments within each group and for the total population yield a higher mean score for the music and sound effects treatment than the narration only treatment. Again, the total means for the high reading group was highest and the low reading group lowest.

In order to determine if the observed differences between the treatments and between the groups were significantly different, the experimental data presented in Table I and Table II were analyzed by two factor analysis of variance with repeated measures. The results are shown in Table III and Table IV.

The results of the two factor analysis of variance with repeated measures for listening comprehension is presented in Table III. The analysis of effects between the reading groups with two degrees of freedom yielded a sum square of 5242.9455, a mean square of 2621.47275, and F Ratio of 49.35601, and a P value of 0.0000 which means the difference is significant at the .01 level. The analysis of the treatments with 1 degree of freedom yielded a sum of squares and mean squares total of 127.73243, an F Ratio of 13.56032, and a P value of 0.00037 which is significant at the .01 level also. The analysis for treatment interaction with 2 degrees of freedom yielded a sum of squares score of 34.00209, a mean square of 17.00104, and an F Ratio of 1.80486, and a P value of 0.16950 which is not significant at the .05 level.

TABLE III

TWO FACTOR ANALYSIS OF VARIANCE WITH REPEATED MEASURES OF THE EFFECTS THAT MUSIC AND SOUND EFFECTS AND NO MUSIC AND SOUND EFFECTS HAVE ON LISTENING COMPREHENSION

Source	DF	Sum of Squares	Mean Squares	F Ratio	P
Between Reading Groups	2	5242 94550	2621 47275	49.35601*	0.00000
Treatment Effects	1	127.73243	127.73243	13.56032*	0.00037
Treatment Interaction	2	34.00209	17.00104	1.80486	0.16950
Within Cell Error	104	979.63595	9.41958		
Total	109	6384.31597			

* Statistically significant at P 0.05

TABLE IV

TWO FACTOR ANALYSIS OF VARIANCE WITH REPEATED MEASURES OF THE EFFECTS THAT MUSIC AND SOUND EFFECTS AND NO MUSIC AND SOUND EFFECTS HAVE ON RECALL

Source	DF	Sum of Squares	Mean Squares	_F_ Ratio	
Between Reading Groups	2	1819.95141	909.97570	46.80713*	0.00000
Treatment Effects	1	62.32917	62.32917	14.25180*	0.00027
Treatment Interaction	2	1.26113	0.63056	0.14418	0.86590
Within Cell Error	104	454.83627	4.37343		
Total	109	2338.37798			

* Statistically significant at P 0.01

The results of the two factor analysis of variance with repeated measures test for recall is given in Table IV. The analysis of effects between the reading groups with two degrees of freedom yielded a sum of squares of 1819.95141, a mean squares of 909.97570, an _F_ Ratio of 46.80713 and a P value of 0.0000 which is significant at the .01 level. The analysis of the treatments with one degree of freedom yielded a sum of squares score of 62.32917, a mean square score of

62.32917, an F Ratio of 14.25180, and a P value of 0.00027 which is significant at the .01 level. The test for interaction with two degrees of freedom yielded a sum of squares score of 1.26113, a mean squares score of 0.63056, and F Ratio of 0.14418, and a P value of 0.86590 which means interaction is not significant at the .05 level.

To determine how each reading level reacted to the separate treatments a single factor analysis of variance with

TABLE V

ANALYSIS OF VARIANCE OF THE EFFECTS OF MUSIC AND SOUND EFFECTS AND NO MUSIC AND SOUND EFFECTS HAVE ON THE LISTENING COMPREHENSION OF HIGH, MEDIUM, AND LOW READING GROUPS

Source	DF	Sum of Squares	Mean Squares	F Ratio	P
High Reading Group	1	4.01389	4.01389	1.08495	0.30470
Medium Reading Group	1	58.86486	58.86486	5.11701*	0.02983
Low Reading Group	1	96.48529	96.48529	7.30254*	0.01079

* Statistically significant at P 0.05

repeated measures was carried out for each reading level for the listening comprehension tests and the recall test. The results are given in Table V and Table VI.

The data presented in Table V is for the independent analysis for the three reading levels of listening comprehension. The high reading group sum of squares and mean square is 4.01389, the *F* Ratio is 1.08495, and the P value is 0.30470 which is not significant at the .05 level. The medium reading group has a sum of squares and mean square score of 58.86486, an *F* Ratio of 5.11701, and a P value of 0.02938 which is significant at the .05 level. The low reading group has a sum of squares and mean square score of 96.48529, an *F* Ratio of 7.30254, and a P value of 0.01079 which is significant at the .05 level.

The data presented in Table VI is for the independent analysis for the three reading levels on the recall test. The high reading group sum of squares and mean square total is 17.01389, an *F* Ratio of 5.54012, and a P value of 0.02433 which is significant at the .05 level. The medium level received a sum of squares and mean square score of 31.13514, an *F* Ratio of 3.23262, and a P value of 0.08134 which is not significant at the .05 level.

The Scheffe F values for comparisons among reading groups for listening comprehension and recall tests are given in Table VII. The comparison of group means for the

TABLE VI

ANALYSIS OF VARIANCE OF THE EFFECTS OF MUSIC AND SOUND EFFECTS AND NO MUSIC AND SOUND EFFECTS HAVE ON THE RECALL OF HIGH, MEDIUM, AND LOW READING GROUPS

Source	DF	Sum of Squares	Mean Squares	F Ratio	P
High Reading Group	1	17.01389	17.01389	5.54012*	0.02433
Medium Reading Group	1	31.13514	31.13514	6.09613*	0.01843
Low Reading Group	1	16.01471	16.01471	3.23262	0.08134

* Statistically significant at P 0.05

high and medium reading groups gives a Scheffe's F test value of 5.9053 for the listening comprehension test and a 6.0632 value for the recall test; both were significant at the .05 level. The comparison of medium and low reading groups gives an F value of 20.3004 for listening comprehension and 18.5765 for recall; both of which are significant at the .05 level. The comparison of the high and low reading groups gives an F value of 46.9881 for listening comprehension and 44.7846 for recall; both again are significant at the .05 level.

TABLE VII

SCHEFFE'S F VALUES FOR COMPARISONS AMONG READING GROUPS FOR LISTENING COMPREHENSION AND RECALL TESTS

Group Comparison	F Value	
	Listening Comprehension	Recall
High-Medium	5.9053 *	6.0632 *
Medium-Low	20.3004 *	18.5765 *
High-Low	46.9881 *	44.7846 *

* Statistically significant at P 0.05

SUMMARY OF FINDINGS, CONCLUSIONS, IMPLICATIONS AND RECOMMENDATIONS

Findings

1. It was hypothesized that there would be no significant difference in the listening comprehension scores of fourth grade students who listened to narrated stories with music and sound effects and those who listened to the same stories without music and sound effects. The analysis of the data reveals that the scores on the listening comprehension tests were significantly higher for those students who listened to the stories with music and sound effects.

2. It was hypothesized that there would be no significant difference in the listening comprehension scores of fourth grade students with low reading abilities who listened to stories recorded with music and sound effects and stories without music and sound effects. The analysis of the data revealsthat the scores on the listening comprehension tests were significantly higher for those students who listened to the stories with music and sound effects.

3. It was hypothesized that there would be no significant difference in the listening comprehension scores of fourth grade students with medium reading abilities who listened to stories recorded with music and sound effects and stories without music and sound effects. The analysis of the data reveal that the scores on the listening comprehension tests were significantly higher for those students who listened to the stories with music and sound effects.

4. It was hypothesized that there would be no significant difference in the listening comprehension scores of fourth grade students with high reading abilities who listened to stories recorded with music and sound effects and stories without music and sound effects. The analysis of the data revealsthat the scores on the listening comprehension tests were higher for those students who listened to the stories with music and sound effects, but the difference was not significant at the .05 level.

5. It was hypothesized that there would be no significant difference in the retention scores of fourth grade students who listened to narrated stories with music and sound effects and those who listened to the stories without music and sound effects. The analysis of the data reveals that the scores on the recall test were significantly higher for those students who listened to the stories with music and sound effects.

6. It was hypothesized that there would be no significant difference in the retention scores of fourth grade students with low reading abilities who listened to narrated stories with music and sound effects and those who listened to the stories without music and sound effects. The analysis of the data reveals that the scores on the recall test were higher for those students who listened to the stories with music and sound effects, but the difference was not significant at the .05 level.

7. It was hypothesized that there would be no significant difference in the retention scores of fourth grade students with medium reading abilities who listened to narrated stories with music and sound effects and those who listened to the stories without music and sound effects. The analysis of the data reveals that the scores on the recall test were significantly higher for those students who listened to the stories with music and sound effects.

8. It was hypothesized that there would be no significant difference in the retention scores of fourth grade students with high reading level abilities who listened to narrated stories with music and sound effects and those who listened to the stories without music and sound effects. The analysis of the data reveal that the scores on the recall test were significantly higher for those students who listened to the stories with music and sound effects.

Conclusions

Based on the data presented in this study and the limitations stated, the following conclusions have been drawn:

1. The addition of music and sound effects to recorded presentations increases the listening comprehension of fourth grade students.

2. Information from a recorded presentation with music and sound effects can be remembered longer by fourth grade students than the same information presented without music and sound effects.

3. Fourth grade students with high, medium, or low reading abilities comprehend and recall recorded information presented with music and sound effects to a greater degree than when the information is presented without music and sound effects.

4. Although the scores for the high reading group were not significantly higher for listening comprehension, for those listening to the music and sound effects stories, their mean scores were higher. After two weeks, those who listened to the music and sound effects did score significantly higher. The research indicates students who are better readers are also better listeners, so a significant difference might not appear immediately after a listening exercise but emerge after a period of time, thus explaining the significantly higher recall scores for the music and sound effects group.

5. The low reading group reacted completely opposite the high reading group by showing a significant difference in favor of music and sound effects immediately after listening to the stories but failed to score significantly higher on the recall test. The low readers, as the research also indicates, are poorer listeners and even though the music and sound effects may have made an immediate difference in comprehension, this difference was not great enough to make a significant difference after a two week delay in recall of information.

Implications

The findings of this study indicate the inclusion of music and sound effects in taped programs for fourth grade students increases their ability to recall information presented. Both long and short term memory of the material is increased. Although the data is not conclusive for all levels, there is a strong indication that the addition of music and sound effects to recorded material helps students at all levels of their language arts development.

Why the addition of music and sound effects should cause such an increase in comprehension and retention is not totally understood, but this and other research indicate several possibilities. The music and sound effects could aid in drawing attention to specific information being presented. The music and sound effects may mask out other possible distractions that may be occuring in the listening environment and allow the student to pay closer attention. The music and sound effects may help creat a mental image of what is being presented. It may be that music and sound effects play an important part in helping to visualize the message by adding "color" to the recorded message.

When students in this study were asked which programs they would like to listen to a second time, they overwhelmingly requested the stories they had heard with music and sound effects. What some educators and radio specialists have been saying for decades seems to be supported by research: music and sound effects added to audio programs do make a difference.

It seems fair to assume that today's children have grown accustomed to audio material being presented in a professional, well produced format and tend to listen more effectively to programs produced in a more sophisticated manner. Because children are bombarded by sights and sounds from numerous sources, they may have become very selective in directing their attention. Sound sources that do not challenge or excite them tend to be disregarded, while sound sources that are exciting to their ears are listened to more concertedly.

The implications of this study should serve as somewhat of a warning for teachers and other educators--that care should be taken in selecting and preparing audio materials for use with children. When children have tape recorders at home with which to play, there is little interest in the technology itself to maintain interest in listening to a boring tape, especially when it is competing with eight tracks and stereo cassettes with all types of stimulating sounds. Besides being interested in the content of a recorded program, the quality of the production itself must be investigated by those selecting materials for children. If the teacher is not selective, the students will be.

The educator does not have to face the full brunt of the selection task alone. Producers of educational materials must share responsibility for the materials they produce. As funds for educational materials become more difficult to secure, the quality of those materials purchased must provide the best quality for the money. As accountability for what educators teach and children learn continues to be stressed, the producers of educational materials must be held more accountable also.

Recommendations

While the findings of this study support the claim that music and sound effects increase the ability of fourth grade students to listen and comprehend information presented on recordings more effectively, several questions remain to be answered. This study dealt mainly with the recall of specific facts from the stories; higher level mental skills such as inference and critical thinking were not tested. Future research should include:

1. studies to investigate the effect of music and sound effects on the inference and critical thinking abilities of students;

2. studies to investigate the possibility that changes in the type of music and sound effects can change perception of what is happening in the story or recorded message;

3. studies to investigate the effect on attitude and possible attitudinal changes resulting in the selection of the music and sound effects;

4. studies to investigate the possible masking effect of over-use of music and sound effects.

Bibliography

1. Adams, Harlen, "Audio-Visual Aids for the Language Arts," Elementary English Review, v. 20, Nov., 1943, 257-264.

2. Adams, Harlen, "Teaching the Art of Listening," The Nation's Schools, v. 34, no. 5, Nov., 1944, 51-54.

3. Blank, M.; Frank, s., "Story Recall in Kindergarten Children: Effect of Method of Presentation on Psycholinguistic Performance," Child Development, v. 42, No. 1, 1971, 299-312.

4. Bohannon, J. N.; and Friedlander, B. X., "The Effect of Intonation on Syntax Recognition in Elementary School Children," Child Development, v. 44, 1973, 675-677.

5. Bormouth, J. R.; Manning, J.; Car, J.; Pearson, D., "Children's Comprehension of Between and Within Sentence Syntax Structures," Journal of Educational Psychology, v. 61, 1970, 349.

6. Devine, Thomas, "Listening," Review of Educational Research, v. 37, No. 2, April 1967, 152-158.

7. Duker, Sam, Listening Bibliography, New York: The Scarecrow Press, Inc., 1969.

8. Friedlander, B. Z.; Cohen de Lara, H., "Receptive Language Anomaly and Language/Reading Dysfunction in 'Normal' Primary Grade School Children," Psychology in the Schools, v. 10, No. 1, 1973, 12-18.

9. Hanpleman, Richard S., "Comparison of Listening and Reading Comprehension Ability of Fourth and Sixth Grade Pupils," Elementary English, v. 35, Jan., 1958, 49-53.

10. Johnson, N. W., "The Elementary School Tunes In," American School Board Journal, v. 104, May, 1942, 28-30.

11. Nichols, Ralph G., "What Can Be Done About Listening?," Supervisor's Notebook, Scott, Foresman Service Bulletin, v. 22, Spring, 1960, 1-4.

12. Petty, Walter, "Listening: Directions for Research," *Elementary English*, v. 39, Oct., 1962, 574-577.

13. Ray, Anne, in Angela Broening (ed.), "Conducting Experiences in English," Monograph No. 8, (Chicago): *The National Council of Teachers of English*, 1939, 202.

14. Rileigh, K. K., "Children's Selective Listening to Stories: Familiarity Effects Involving Vocabulary, Syntax, and Intonation," *Psychological Reports*, v. 33, 1973, 255-266.

15. Russell, David H., "A Conspectus of Recent Research on Listening Ability," *Elementary English*, v. 41, March, 1964, 262-267.

16. Smart, Mollie S.; Smart, Russell C., *Children: Development and Relationships*, New York: Macmillan Publishing Co., 1977.

17. Snyder, R.; Pope, P., "Auditory and Visual Inadequacies in Maturation at First Grade Levels," *Journal of Learning Disabilities*, v. 5, 1972, 620-625.

18. Thompson, R. A., "Developing Listening Skills to Improve Reading," Education, v. 91, No. 3, 1971, 261-269.

19. Turkewitz, G.; Birch, H. G.; Cooper, K. K., "Patterns of Response to Different Auditory Stimuli in the Human Newborn," *Developmental Medicine and Child Neurology*, v. 14, 1972, 487-491.

20. Wetstone, Harriet S.; Friedlander, Bernard Z., "The Effect of Live, TV, and Audio Story Narration on Primary Grade Children's Listening Comprehension," *Journal of Educational Research*, v. 68, No. 1, Sept. 1974, 32-35.

CHILDREN'S UNDERSTANDING OF IMPLIED MOTION CUES

By

RONALD A. SAIET, Ed.D.

Associate Director for
Learning Services,
Northeastern Illinois
University, Chicago

Prepared for presentation at Association for Educational
Communications and Technology
1979 National Convention
New Orleans, Louisiana

ABSTRACT

Children's Understanding of Implied Motion Cues

Ronald A. Saiet

Extrapolating from Piaget's theory of cognitive development it can be argued that graphic conventions such as flow lines and vibration marks used by artists to depict motion would have little or no meaning to pre-school children, but that an action pose, being more like real human posture, would be a symbol within the grasp of understanding of children in this group. By age seven, most children should be capable of understanding arbitrary graphic symbols, and by age eleven virtually all should be capable.

The kind of object upon which a motion cue is said to act might also affect perception of movement. Human figures, for example, might appear to be more likely to be in motion than abstracted geometric shapes. An inherent motion factor may also be said to reside within an object, such that it psychologically appears to be more active or more static--running figures, ellipses, and rectangles may be seen as more active than standing figures, circles or squares.

An experimental study was conducted in an attempt to determine how children --four, seven, and eleven years old--perceive implied motion cues in combination with a variety of objects. Ninety children, 30 in each age group, were shown 24 pictures. There were three kinds of objects, each with an inherent active and static dimension. Each of these six pictures appeared with one of three motion cues, and alone. The number of each of the picture types identified as moving by each subject was determined. An age x object x activity of object x motion cues repeated measures ANOVA was used to evaluate the data.

All main effects and all interactions were significant. It was concluded that children at different stages of development differ in their interpretation of certain graphic conventions. Illustrators, editors, and publishers should take into account the pictorial literacy of an intended audience when they design materials to convey motion information.

"Of all the senses, trust only the sense of sight."
Aristotle, Metaphysics

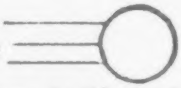
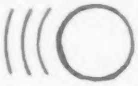
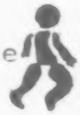
Little more than a decade ago, Fleming (1966) observed that in the midst of new trends emerging in educational technology, the commonplace print media continued largely unchallenged as the most widely used educational resource (p. 299). He also found a proliferation of illustrations in textbooks, one illustration appearing every 1.58 pages of 40 eighth grade texts sampled (p. 317). Casual observation of textbook offerings in children's literature today reveals that the situation now appears to be much as it was then, and one continues to wonder along with Fleming what essential role illustrations serve; they add to the size and costs of printed materials, but is there any evidence that pictures add to a child's understanding of the matter before him?

Gropper (1963) observed that much less is known about how people respond to visuals than about how people respond to words. Art students are taught what Dondis (1973) called the basic elements of visual communication--dot, line, shape, direction, tone, color, texture, dimension, scale, and movement (pp. 39-66)--but Hengen (1970) asked whether children have acquired an understanding of these elements, and if so, whether they are acquired through talent or through training (p. 3). The present writer's survey of the literature has found it wanting in questions and answers about how people--children and adults--perceive and learn from non-verbal, non-moving visual stimuli.

The problem for message designers was stated by Fleming as follows: "It is improbable that an instructional system can approximate optimum performance unless the designer of that system is fully informed on the relevant characteristics of the system component options. Pictorial stimuli will be one of the system component options the designer will select to meet certain system specifications" (1966, pp. 317-318).

It is the aim of this paper to examine a small part of this problem within the area of pictorial communication. The investigation carried out here was suggested by a doctoral dissertation by Satterthwaite (1965), "An investigation of selected cues for implying motion in non-motion media." It was Satterthwaite's assumption that graphic devices (called "implied motion cues") commonly used in two-dimensional print media by artists and illustrators sucessfully transmit some motion information to the perceivers of these graphic devices. He then selected and manipulated three such implied motion cues*, and asked "Do all of the possible variations of these implied motion cues transmit the same motion information or are some combinations of cues and objects best suited for the transmission of certain types of motion information?" (pp. 20-25) Using 40 subjects, all students enrolled in a graduate course in Audio-Visual Communications at Indiana University, Satterthwaite's basic

---

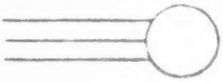
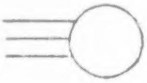
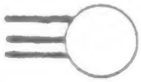
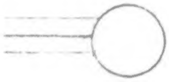
* 1. Flow line  2. Action line  3. Action pose 
(See page 7 for definitions.)

assumption that the cues successfully implied motion was upheld and the answer to his basic question was that certain variations and combinations of cues and objects did affect the communication differentially (pp. 123-126).

It is not surprising, at least to this investigator, that students in a graduate audio-visual course understood the intention of the implied motion cues. They especially would seem likely to have encountered such cues in their lifetimes and to have assigned accurate meanings to them (accurate because their interpretation was congruent with the designer's intended meaning). It is not surprising, either, that even this group of subjects did not all interpret the variations of the cues in the same way, because most of Satterthwaite's variations (i.e.; long/short, thick/thin, and others*) do not have universal meaning, were never systematically taught or learned by the subjects, and are arbitrary in nature. As Knowlton (1966) observed, a sign's ability to signify a concept is dependent upon the concept's having already been attained by its interpreter (p. 165). Fat and thin lines intended to imply more or less speed, for example, are arbitrary signs within this context, and have to be learned if other than idiosyncratic interpretations are to be given.

---

*

1. long flow line 
2. short flow line 
3. thick flow line 
4. thin flow line 

People are not born literate. Through a combination of factors--development, experience, formal education--people eventually learn to decode visual messages. The basic alternative question to Satterthwaite's which is to be posed then, is when do graphic devices effectively transmit motion information? Put another way, at what stage of development do children accurately perceive visual cues designed to imply motion from non-moving pictorial stimuli?

According to Piaget's theory of cognitive development one either assimilates new information into an already existing mental category (psychological structure) as something like or unlike something already understood, or one accommodates and changes his mental structure--creates a new category--to deal with information which cannot be so assimilated. Furthermore, each child proceeds through a series of stages, each stage represented by the presence of particular psychological structures. Since experience is always changing, new structures are always evolving, with the result that adult structures are more sophisticated than those available to children. Highly evolved adult intelligence is represented by more assimilation into existing, well developed structures whereas learners at earlier stages of development, not having yet evolved the necessary structures present at later stages of cognitive development, are prone to misinterpret (or ignore) information.

If a child perceives a pictorial element as familiar, he will be able to process it, to make sense of it. If on the other hand, he is unable to recognize a pictorial element--person, place, thing, or action--this might be construed as an indication of his picture illiteracy. An implied motion cue, if not within the range of a child's experience, would be incomprehensible to him. A viewer at a higher level of picture literacy, possessing only partial literacy, would make sense of some parts of pictures but would misinterpret the meaning of the entire pictorial message; if a single cue is misinterpreted, meaning is incomplete. Attained picture literacy would be evidenced by one's ability to correctly interpret complex visual symbols, abstractions, and graphic conventions.

Considering the expense publishers incur in creating illustrative materials, it would seem to be useful to them to know at what stages of development their audiences are--their level of picture literacy--so that the pictorial information can be made to be appropriate to each audience. One might assume that an American child, even at age four, will recognize pictorial representations of a familiar object because of his experience with television, film, and books. But it is another matter to predict with any accuracy at what age he will be able to infer motion from two-dimensional motion conventions. It seems plausible, however, that as children get older they recognize the intended meaning of certain cues; that certain motion cues will be perceived more often as containing motion information by children as they mature; and that inherent characteristics (such as shape and life-likeness) of objects purported to be in motion affect the communicative power of motion cues as perceived by children at different stages of development.

The study undertaken by the current writer investigates the problem of when and how children at different stages of development understand implied motion cues. The current vehicle of presentation precludes a more thorough review of the relevant, albeit limited, literature. For additional explication of Piaget's theory relevant to the current exloration, see Ginsberg & Opper (1969), Phillips (1975), Rohwer, Ammon & Cramer (1974), and Wadsworth (1971). Research conducted by the High/Scope Educational Research Foundation (Banet, et. al., and Nederveld & Thomson, 1972) into children's undertanding of representative material and the implications for learning is useful, and Friedman & Stevenson's study of developmental changes and their affect on understanding of implied motions cues (1975) provides an alternative experimental approach utilizing an interesting array of stimulus materials, and was most useful in developing the current study.

Problem and Methodology

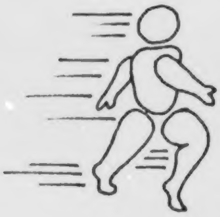
Problem

This study was designed to find out how children--4, 7, and 11 years old--perceive selected implied motion cues. The implied motion cues selected for this study were the flow line, the vibration mark, and the blur. The drawings used were selected by three judges (graphic artists) according to two criteria: they had to be similar to the postural characters and cartoon devices used in Friedman and Stevenson (1975), and they had to convey motion information to them (i.e., to the judges).

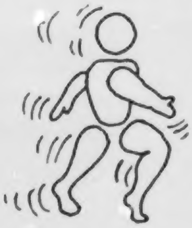
Definitions

Implied motion cues...those graphic devices commonly used to represent motion in non-motion media (i.e., children's books, textbooks, magazines, and newspapers).

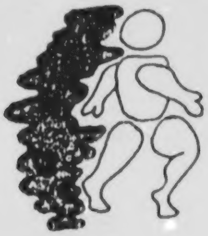
Flow line............a graphic device used to imply speed, it consists of even, clearly spaced, parallel lines that extend horizontally from the trailing edge of the object purported to be in motion.



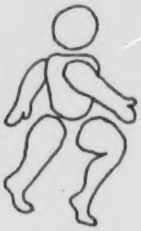
Vibration mark........a graphic device used to imply erratic movement, it consists of short lines which echo a part of the figure purported to be in motion.



Blur..................a graphic device used to indicate great speed, it consists of a shaded area extending from the trailing edge of the figure purported to be in motion.



Action pose..........Satterthwaite defines the action pose as an implied motion cue "which is an integral part of the object itself" (1965, p. 18). It consists of a representation of a human figure in profile with its body and limbs arranged in a manner deviating from a strictly vertical, standing position. For the purposes of statistical treatment in this study, however, the action pose is regarded as an active object, not as a motion cue per se.



The objective of this study was to measure motion perception of subjects in three age groups (four, seven, and eleven year olds), corresponding to Piaget's pre-operational through concrete operational periods, to determine whether (a) the implied motion cues are perceived, whether (b) there is an hierarchical order of identification of the selected cues, and whether (c) the configuration of the object purported to be seen as moving affects the perception of motion. The configuration of object type may be said to contain two dimensions, shape and activity.

The questions, then, which guided this study were:

1. Do children at different developmental levels attend to implied motion cues, and do they read them as motion devices?

2. Is one selected implied motion cue perceived more often as containing motion information than is another selected implied motion cue by children as they advance from the pre-operational stage to the concrete operations stage?

3. Do certain inherent characteristics of the kind of objects being acted upon differentially affect the communicative power of selected implied motion cues as reported by children as they mature from the pre-operational to the concrete operations stage?

The last of these questions requires some additional consideration. Both Satterthwaite, and Friedman and Stevenson (1975) found the nature and configuration of the object critical to the question of motion perception in non-motion media. Each study found that active, postural figures were full of motion information for all of their subjects, regardless of the conventional cues applied. Friedman and Stevenson did not, however, include geometric figures in their study. Satterthwaite did, but found no perceivable differences in the speed of active or static geometric figures, although rounded objects were seen as moving faster than angular objects.

This study attempted to examine all three kinds of objects--human figures (postural), angular geometric, and rounded geometric--to assess their interaction with the motion cues selected for this study and with subjects at their various stages of development. Additionally, an active and static object were presented for each kind of object. The objects may be classified as in Figure 1.

	Active	Static	Human figure	Geometric, angular	Geometric, rounded
Running figure	X		X		
Standing figure		X	X		
Rectangle	X			X	
Square		X		X	
Ellipse	X				X
Circle		X			X

Figure 1. Classification of kinds of objects.

Methodology

Subjects. In this study there were 30 four year olds, 30 seven year olds, and 30 eleven year olds. They were selected arbitrarily from the National College of Education Baker Demonstration School, and the Martin Luther King, Jr., Laboratory School in Evanston, Illinois. The subjects were primarily from middle to upper-middle class families.

Materials. A series of 24 cards served as the stimulus materials. Each card contained one illustration of one of six objects: circle, ellipse, square, rectangle, running figure, and standing figure. The objects were of three kinds: human figures, angular geometric figures, and rounded geometric figures. There were two dimensions of each object type: active and static. The active human object was represented by a running figure. The static human object was represented by a standing figure. The active angular geometric object was represented by a rectangle. The static angular geometric object was represented by a square. The active rounded geometric object was represented by an ellipse. The static rounded geometric object was represented by a circle. Three implied motion cues were added to one picture of each object, and one picture of each object stood without any implied motion cue, for a total of 24 pictures. The selected implied motion cues were the flow line, vibration mark, and blur. Figure 2 illustrates the 24 pictures.

Procedures. First, the four year olds were screened to ascertain their grasp of the concept of real movement. They were asked to stand still, to run, to jump, to move a "little bit", to "shiver and shake like when you are cold." They were then given a doll, a ball, a rectangular block, a square block, and an elliptical block and asked if each was moving, and could they make the objects move. Two subjects who did not understand the concept of real movement were replaced.

All subjects were given a classification test which involved labeling each picture as "moving" or "standing still." The exercise was carefully explained to each subject; he was told that if any picture seemed as though it were moving, or as though it were supposed to be moving, even a little bit, it should be classified as moving. The subjects were handed one picture at a time--a different random order was used for each subject--and the experimenter (the author) asked the

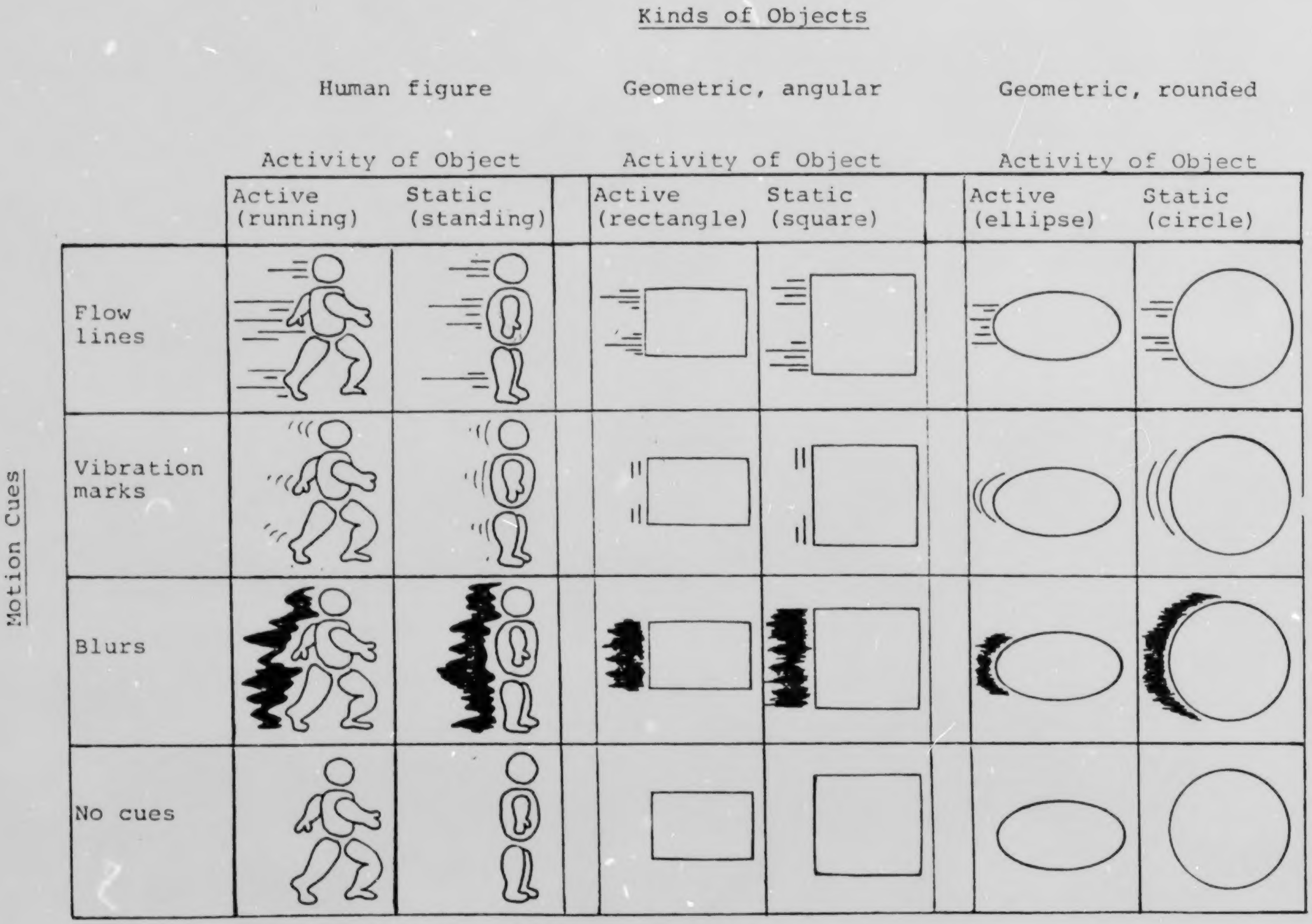
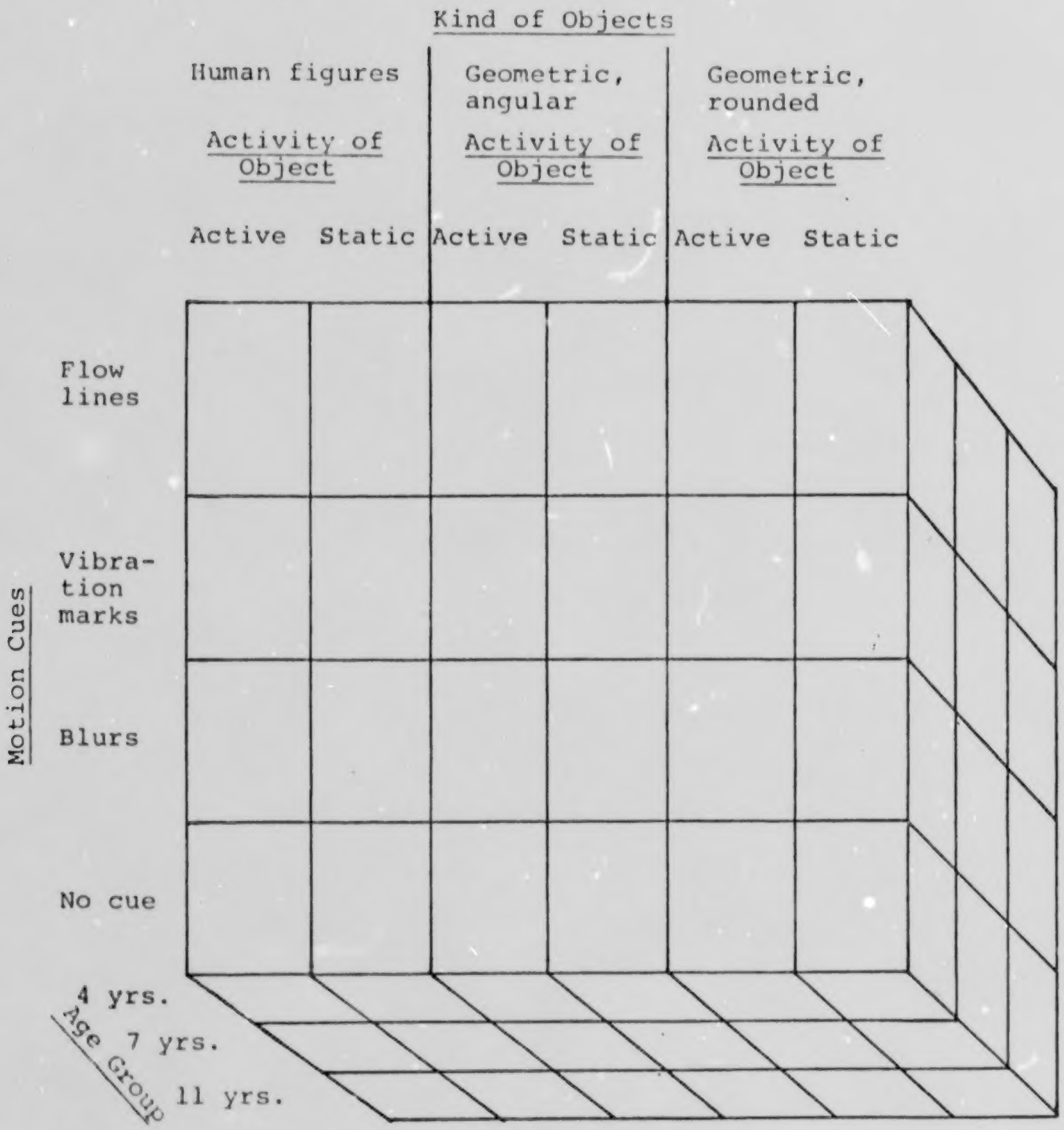

Kinds of Objects
Human figure
Geometric, angular
Geometric, rounded
Activity of Object
Activity of Object
Activity of Object
Active (running)
Static (standing)
Active (rectangle)
Static (square)
Active (ellipse)
Static (circle)
Motion Cues
Flow lines
Vibration marks
Blurs
No cues

Figure 2. Illustration of pictures used in study.

children whether they thought that the thing in the picture looked like it was supposed to be moving or standing still. The experimenter recorded an *M* on the picture if the object was described as moving, or an *S* on the picture if the object was described as standing still.

After the test was completed, the 24 cards were gathered and stapled in the same sequence as presented, and the experimenter interviewed each student. The interview was informal, and students were asked to explain the reasons for their classifcations. The experimenter's expectations were that the older children would identify the graphic conventions, whereas the youngest children would probably pay more attention to the nature of the object. The dialog would provide additional insights into how children read the pictures. (Hengen suggested that extensive interviewing be conducted to facilitate understanding the thought processes involved when a subject interprets an illustration (1970, pp.48-49). It is also the technique used by Piaget.)

Appendix A contains the raw data used in this study. Each object/motion cue picture identified as moving received a score of *1*, each identified as standing still scored *0*. An age group x sex x kind of object x activity of object x motion cues analysis of variance, with repeated measures on the last three variable, was used to analyze and evaluate the data (Winer, 1971). The experimental design is shown in Figure 3.



Kind of Objects
Human figures
Geometric, angular
Geometric, rounded
Activity of Object
Activity of Object
Activity of Object
Active
Static
Active
Static
Active
Static
Motion Cues
Flow lines
Vibration marks
Blurs
No cue
Age Group
4 yrs.
7 yrs.
11 yrs.

Figure 3. Illustration of experimental design.

Results and Discussion

The analysis of variance showed neither the main effect of sex nor its interactions with other variables were significant. To simplify the presentation of the results, a second analysis was done from which the variable sex was excluded. The results of the first analysis will not be discussed. The summary of the second analysis is shown in Table 1. As shown in Table 1, all main effects and interactions were significant.

The primary concern of this study was to see how a child's stage of development is related to his ability to perceive or "read" conventional graphic devices described here as implied motion cues. Additionally, it was speculated that both the kind of object and the object's inherent motion characteristics might be differentially related to the manner in which children at different stages of development interpret intended motion symbols. The four-way age group x kind of object x inherent activity of object x implied motion cue interaction, illustrated in Figure 4, seems to provide the best framework within which to examine the outcomes of this study. All lower level interactions appear to support the observations made based upon the higher order interaction without adding new insights and, therefore, will not be addressed specifically.

Because of the complexities involved in relating the results of the four-way interaction, the discussion which follows will be presented in the following manner. The human figure will be discussed first because of its dramatic contrast to the other kinds of objects in its active dimesion, and because of the

Table 1

Summary of Analysis of Variance

Source	*df*	*MS*	*F*
Age Group (G)	2	26.15	55.86*
Error	87	.46	
Object (O)	2	13.29	87.76*
O x G	4	1.76	11.65*
Error	174	.15	
Activity (A)	1	26.88	260.36*
A x G	2	.87	8.47*
Error	87	.10	
O x A	2	28.58	263.93*
O x A x G	4	.82	8.19*
Error	174	.10	
Motion Cue (C)	3	25.19	122.92*
C x G	6	4.43	21.63*
Error	261	.20	
O x C	6	1.42	17.34*
O x C x G	12	.42	5.11*
Error	522	.08	
A x C	3	.66	11.81*
A x C x G	6	.12	2.19**
Error	261	.05	
O x A x C	6	.37	5.59*
O x A x C x G	12	.20	3.07*
Error	522	.06	

*$p < .001$

**$p < .05$

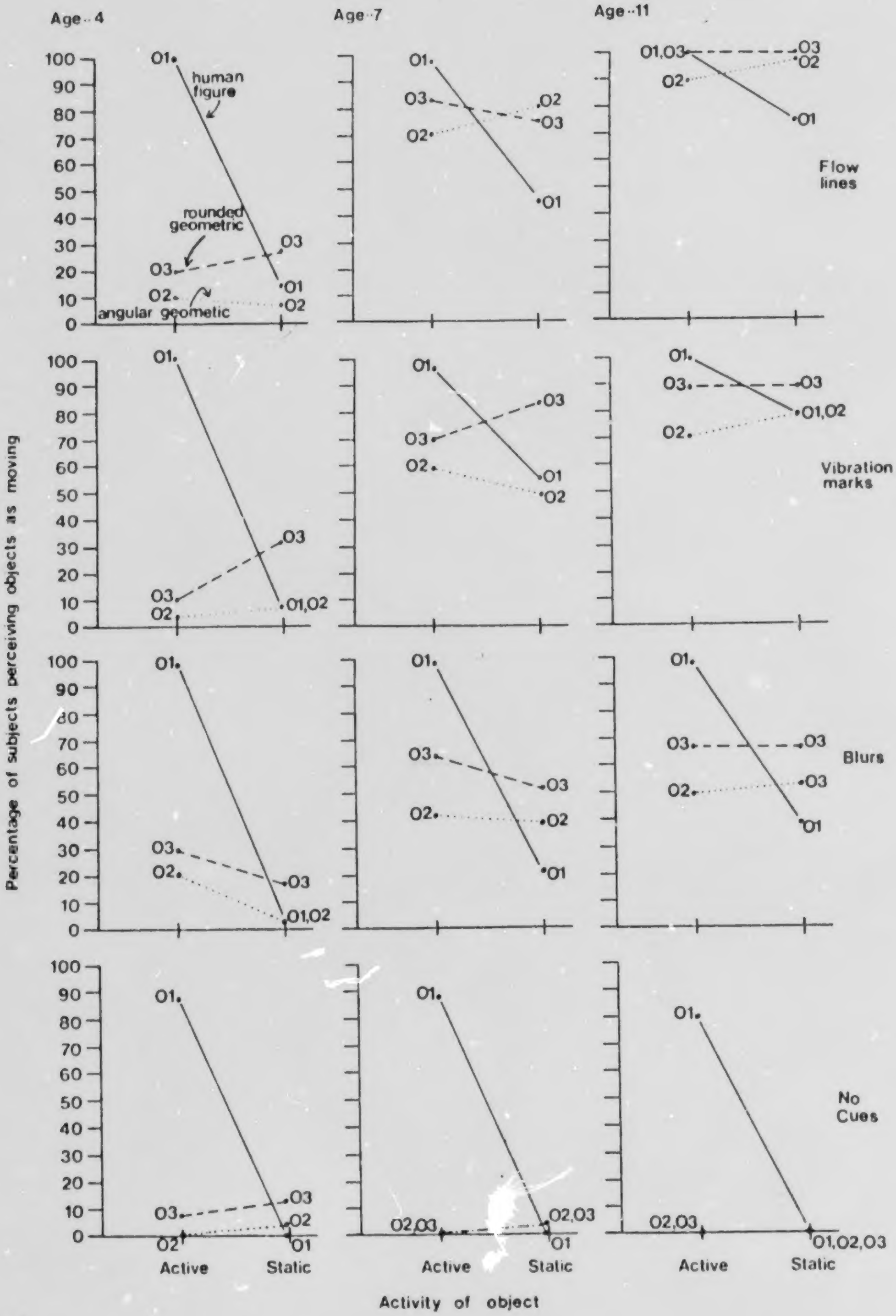


Figure 4. Illustration of the four-way interaction: age group x kind of object x activity of object x motion cue.

variable affect age seemed to have upon reading of motion in its static dimension when motion cues were added. Following that, because developmental level and the reading of motion cues were of primary interest to the study, observations will be made about what has been learned about the age group and motion cue variables. Then observations will be made about the more abstract geometric shapes and their inherent activity--as they relate to each other as well as to other variables. Conclusions will be stated and suggestions for further research will be advanced.

Reading the Illustration of the Four-way Interaction

A guide to reading the four-way interaction is presented in Figure 5. The reader is advised to examine that figure before exploring the illustration of the four-way interaction in Figure 4.

Reading the graphs in Figure 18 from left to right demonstrates the different ways in which the three age groups reported perceived motion for a given object type and activity state. Reading from bottom to top shows how the addition of a particular motion cue may have altered those perceptions.

Human Figure

One dramatic result was that the active (running) human figure was almost always seen as moving, regardless of either age of subject or motion cue applied. The running figure, or action pose, clearly possesses inherent motion information for all age groups studied.

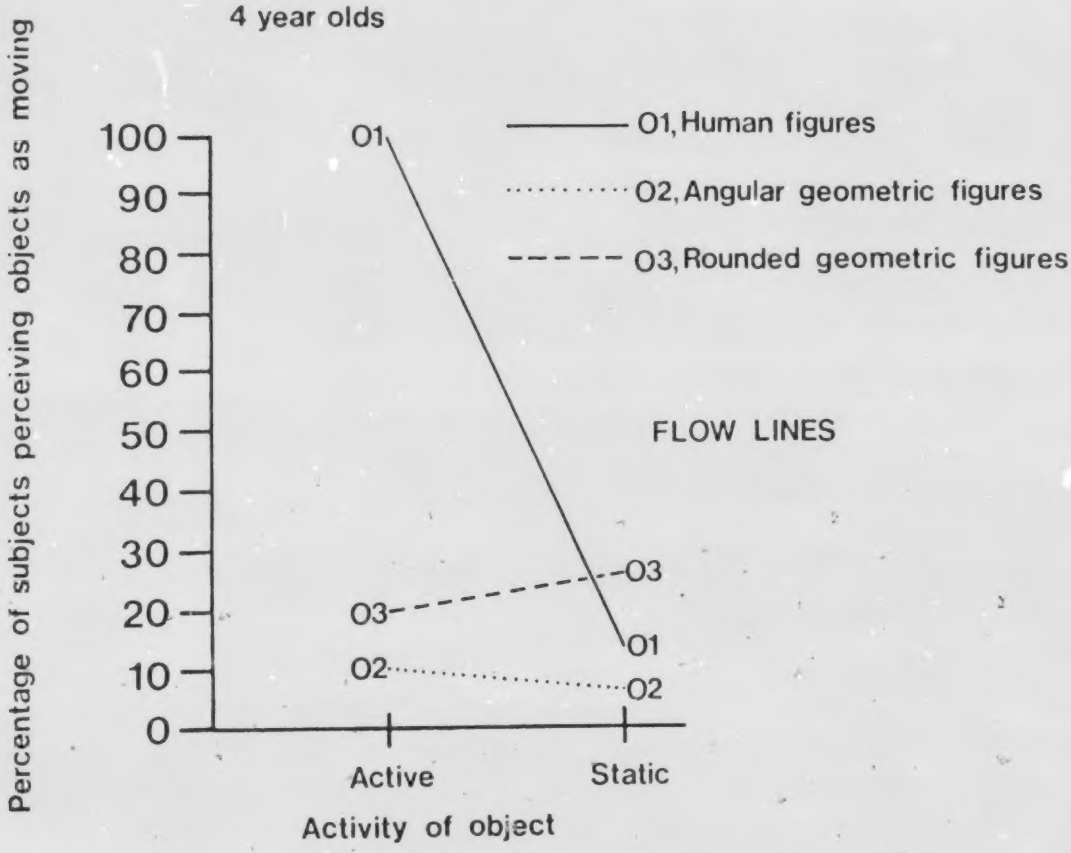


Figure 5. Guide to reading illustration of four-way interaction. The graph in this figure is the same as the graph in the upper left corner of Figure 4. It shows the percentage of 4 year olds who perceived pictures as moving when flow lines were added to each of three objects, for either of two activity states. The solid black line shows that, in the presence of flow lines, 100% of the 4 year olds saw the active (running) human figure as moving, while 13% saw the static (standing) human figure as moving. The dotted line shows that, in the presence of flow lines, 10% of the 4 year olds saw the active (rectangle) angular geometric figure as moving, while 6% saw the static (square) angular geometric figure as moving. The dash-line shows that, in the presence of flow lines, 20% of the 4 year olds saw the active (ellipse) rounded geometric figure as moving, while 26% saw the static (circle) rounded geometric figure as moving.

The static (standing) human figure clearly presents quite different information. Although it is almost consistently reported as having less motion information than the other static figures--square and rectangle--the addition of motion cues apparently differentially affects the amount of motion perceived by children at different stages. All three age groups reported no motion for the static figure under the no cue condition. As any one of the kinds of cues was added, however, the two older groups reported that the standing figure was "moving." The pre-schoolers, on the other hand, reported almost no motion.

This latter result seems reasonable. It would be totally inconsistent with reality and most experience for yound children to assume that a standing human figure, feet together, body rigid, was moving. The presence of abstract motion cues apparently did not override "no-motion" signals inherent in the standing figure. The fact that in most cases the older children reported less motion for the static human figure than for the static geometrics (although the statistical significance of this was not tested) may be an indication that even these children were a bit reluctant to identify a "standing" figure as "moving."

Age Group and Motion Cue

In general, the four-way interaction seems to reflect the fact that with increasing age the various motion cues had an increasingly heterogeneous effect on the perception of motion as the motion cues added were, respectively, blurs, vibration marks, and flow lines. The percentage of subjects perceiving the five objects (other than the running figure) as moving in the graphs in Figure 4 for the seven and eleven year olds increases with respect to that same order of addition of motion cues, while the four year olds' graphs remain essentially unchanged from the no cue condition. An explanation consistent with this observation would be that the older children were "reading" the cues, while the pre-schoolers were ignoring them, centering their attention on the object depicted.

In interview conducted by the author with the students following the administration of the experiment, the older groups frequently explained that things seemed to be moving because of "the lines", or motion cues. Only one four year old identified the lines at all. That the older children were probably "reading" the cues is also suggested by the differential effect the various cues apparently had. For the seven and eleven year olds flow lines and vibration marks seemed to be more effective than the blur motion cue for most objects. This is plausible because the interviews revealed that the blur was probably a more ambiguous cue than the other graphic conventions. It received the widest range of interpretations, several inconsistent with motion interpretation. Some saw it appropriately as a "bunch of little lines close together, really moving fast." But others saw the blur as a "tornado", or "air pushing", or merely as "a wall", or "a shadow." Perhaps in the absence of contextual clues the blur motion cue as drawn for this study was a more confusing sign than were the other two graphic cues. The fact, however, that even a poorly drawn cue was reported on the average more than 50 per cent of the time as contributing to motion interpretation may be further evidence that pictorial literacy has advanced enough by the age of seven to differentiate the older groups from the nursery school children.

Object and Activity

The power of the running human figure as a motion indicator has already been discussed. A closer look at the percentage of subjects perceiving movement in the geometric figures suggests that the variance between the two kinds of geometric shapes could be chane variation. In all but one instance (see the top center graph in Figure 4) rounded geometrics were perceived as moving more than angular geometrics, regardless of activity state inherent in the objects; the almost horizontal slopes and sometimes nearly parallel nature of the lines in Figure 4 suggests that differences between classes of geometric objects were not very meaningful.

Similarly, the percentage of subjects perceiving movement in the geometric figures suggests that the variance within the activity dimension could also be chance variation in most cases. This is almost surely the case regarding the active figures--ellipse and rectangle--and usually the case regarding the static figures--circle and square. It is possible, however, that under certain conditions (see, for example, the seven year old/vibration mark graph in Figure 4) the circle may be a better motion indicator than the square, albeit the statistical significance of this was not tested.

In interviews conducted by the experimenter several of the four year olds expressed some ideas about the geometric objects which may explain some of the differences between subjects' responses to geometric objects--both active and static--and to the active human figure. Comments like "a square can't move" were very common among the pre-schoolers. They also said that "the man is running" or "the man is jumping." This may indicate that as four year olds look at the geometric shapes, perhaps some intuitive ideas about friction or inertia are operating.

Conclusions

The results of this study seem consistent with the findings of Friedman and Stevenson (1975). In both studies, younger and older subjects did not classify pictures in the same manner. Reliance on implied motion cues apparently does increase with age.

The current study, using seven year olds rather than six year olds, seems to establish more precisely the time period when children may begin to "read" motion cues. Friedman and Stevenson found that four and six year olds--both within the pre-operational stage--saw most postural pictures as "moving" and that they did not see any difference in the way graphic devices affected motion perception. Sixth graders and college students classified approximately the same number of postural pictures with or without motion cues as "moving" (p. 777). The current study found the behavior of seven and eleven year olds--at the beginning and end of the concrete operations stage--to be more like each other than like the behavior of the pre-operational four year olds. It may be that by the time youngsters enter the concrete operations period they have the ability to "read" conventional motion cues, but not much before. The current study used geometric shapes as well as postural figures; Friedman and Stevenson used only the latter. That the older children seemed to attribute a substantial amount of motion to both kinds of objects when motion cues were used may support and emphasize Friedman and Stevenson's suggestion that once implied motion cues are learned, they assume the same power in eliciting a reading of movement as do action pose kinds of cues.

Comparisons with the Satterthwaite study (1965) must be more conservative. Satterthwaite was concerned with the quality and degree (i.e., smoothness, speed and distance) of motion as reported by college students. Although some similarities seem to exist between the studies--rounded objects were generally seen more often as moving than were angular ones, active and static geometric figures seemed to have similar amounts of motion information, and the running figure was always more potent than the standing figure--tests for simple effects were beyond the scope of either study.

The facts which do seem to emerge clearly from this study are that the active human postural figure was understood by the children in this study, regardless of their age group, as moving with or without the addition of motion cues, and that once the children left the pre-operational stage of development they apparently had the ability to "read" the implied motion cues used in this study. The implication for message designers in general and for those interested in illustrating books for young children in particular seems clear. The differences between children's ability to understand implied motion cues at their respective stages of development ought to be taken into account when producing pictorial materials.

Suggestions for Further Research

Other selected motion cues. Only three implied motion cues were selected for this study, and one of them--the blur--was apparently more ambiguous than the other cues, perhaps because it was poorly drawn. A study using additional motion cues, or variations of the cues used in the present study could be developed. The relative power of different cues to elicit motion interpretations from pictorially literate subjects could be explored.

Contextual clues. All of the pictures used in the current study were essentially context-free and in that sense were unrealistic. It would be interesting to further test Stevenson and Friedman's hypothesis--that pictorial cues with structures very similar to structures in the three-dimensional world would be easy to interpret, but that pictorial cues less than isomorphic with reality would be more difficult to interpret (p. 773)--using more lifelike pictures and settings than were used in the current study. For example, Figure 6 shows a boy on a bicycle.

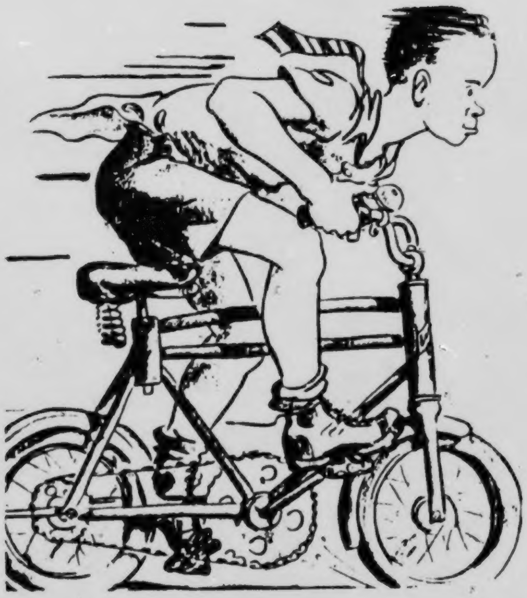


Figure 6. Illustration of Robert McCloskey's cartoon technique. Note the artist's use of flow lines. From Make Way for Ducklings, by Robert McCloskey, copyright (c) 1948. By permission of the Viking Press, Inc.

Most readers would agree that he and his bicycle are supposed to be in motion. The illustrator, Robert McCloskey, has provided many motion cues. Although the cartoon technique, including the use of flow lines, is an abstraction of reality, there are enough lifelike cues such as the boy's position on the bicycle, his tie, shirt-tail, and hair flying to assure that most readers would acknowledge that the boy and the bicycle are "moving." One could, however, manipulate this drawing, maintaining some contextual clues while eliminating others. Would some readers, for example, think that the boy and his bicycle (in Figure 14) were moving if the boy were drawn sitting erect, his back perpendicular to the ground, his feet on the pedals? Would this be the contextual equivalent of the standing figure in the current study?

In the current study, in the absence of contextual clues, it appeared that pre-schoolers ignored, for example, the flow line motion cue. Would a young child be more likely to "read" a flow line if it helped him to interpret the picture within a specific context? Imagine a drawing of a boy in the process of climbing a flagpole. Might the placement of flow lines above or below the boy be used to indicate if he was climbing up or down the pole, and would young children use such cues to decide the direction of motion? Experiments could be designed to explore these and similar questions.

Satterthwaite follow-up. Satterthwaite explored several ideas which the current study did not. It would be interesting to submit to a more rigorous statistical treatment questions about perceived tense (e.g., in the presence of different cues, or of cues of greater length or thickness, does an object appear to be about to start, already in motion, or about to stop?); speed (e.g., does an object in the presence of one implied motion cue appear to be moving faster than the same object in the presence of another implied motion cue?); or quality of motion (e.g., which motion cues are best used to indicate erratic motion; fluid motion?).

Survey of artists. It would be interesting to survey artists to find out what kinds of assumptions, if any, they make about children's pictorial literacy, and to ascertain what techniques they would use under what conditions to imply motion. They might be asked if they attempt to adapt their techniques to various audiences, or if they would modify their style if it were demonstrated that groups of readers responded in different ways to different styles. A survey instrument could be designed to learn more about the community of children's book illustrators, and these and other questions could be asked to learn more about the concerns and attitudes of this group of "message designers."

Summary

This study reports an investigation of how children at different stages of development understand implied motion cues. The function of these cues is to represent motion in non-motion media (i.e., children's books, textbooks, magazines, and newspapers). The implied motion cues selected for this study were the flow line, the vibration mark, and the blur.

Problem

Prior research was examined which had shown that visually encoded messages communicate in a manner similar to verbally encoded ones. How well a person understands the significant verbal symbols of a language and their temporal and spatial arrangement is a measure of one's verbal literacy; how well a person understands a language composed of visual symbols and their spatial arrangement can be said to be an index of one's visual literacy. The present study was interested in what might be called pictorial literacy and more specifically, with respect to a particular element of pictorial communication, namely, implied motion cues.

People are not born literate. Through a combination of factors, including development, general experience, and formal education, people eventually learn to decode written and drawn messages. Extrapolating from Piaget's theory of cognitive development, it was suggested that a parallel exists between the development of spoken language-symbol systems and the development of other-symbol systems, specifically, pictorial motion symbols. According to this view graphic conventions such as flow lines and vibration marks used by artists to depict motion would have little or no meaning to preoperational (pre-school) children, but that an action pose, being more like real human posture, would be a symbol within the grasp of understanding of children within this group. By the beginning of Piaget's concrete operations stage, about seven years of age, most children whould be capable of understanding arbitrary graphic signs, and by age eleven virtually all should be capable.

It was further suggested that the kind of object upon which a motion cue is said to act might also affect perception of movement. Human figures, for example, are more likely to appear to be in motion than are geometric shapes such as squares or circles. It was speculated that an inherent motion factor may also be said to reside within an object, such that it appears to be more active or more static--e.g., running figures, ellipses and rectangles may be seen as more active than standing figures, circles or squares.

Three questions were posed.

1. Do children as they grow older attend to flow lines, vibration marks, and blurs and do they read them as motion devices?

2. Is one selected implied motion cue perceived more often as containing motion information than is another selected implied motion cue by children as they advance from the preoperational stage to the concrete operations stage?

3. Do certain inherent characteristics of the kind of objects being acted upon differentially affect the communicative power of selected implied motion cues as reported by children as they mature from the pre-operational to the concrete operations stage?

Methodology

Subjects. Ninety primarily middle class subjects--30 four year olds, seven year olds, and eleven year olds--were selected arbitrarily from schools in Evanston, Illinois.

Procedures. An experimental study was conducted in an attempt to determine how children at different stages of development perceived implied motion cues in combination with a variety of objects. Each subject was shown the same 24 pictures which were arranged in different random orders for each child. There were three kinds of objects--human figures, angular geometric, and rounded geometric--each with an inherent active and static dimension. Each of these pictures appeared with one of three motion cues and alone. The number of each of the picture types identified as moving by each subject was determined. An age group x kind of object x activity of object x implied motion cue repeated measures analysis of variance was used to evaluate the data.

Results and Conclusion

All main effects and all interactions were significant. The four-way interaction, *F*(12, 522) = 3.07, *p* .001, was illustrated and used as the basis for presenting the results.

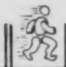
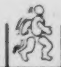
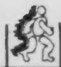
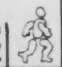
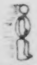
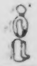
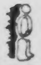
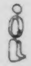
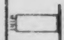
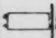
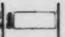
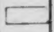
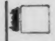
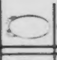
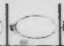
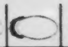
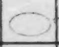
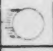
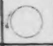
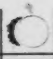
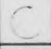
The data seemed to show clearly that the active human postural figure was understood by all children in the study as "moving" with or without the presence of motion cues. The pre-operational children appeared to attend to only the kind of object, the action pose being the only object type for which significant movement was reported. Motion cues seemed to have no effect for this group. The older groups both reported apparently similar motion increases for all object types and activity states when motion cues were added--except that the active human figure remained consistently high. It was not determined whether any motion cue was significantly more effective than any other, although the blur appeared to be a somewhat weaker motion signifier in this study. This result may have been due to the blur's being rather poorly drawn.

It was concluded that once children leave the pre-operational stage they apparently have the ability to "read" the implied motion cues used in this study. The implication for message designers is that this differential ability ought to be taken into account when producing pictorial materials for children.

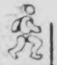
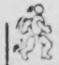
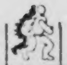
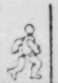
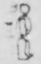
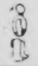
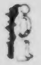
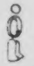
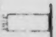
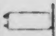
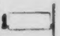
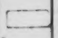
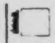
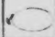
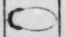
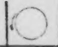
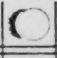
APPENDIX

Raw Data

Raw Data: 4 year olds

Subject No.																									Totals (by subject)
1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
2	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	1	1	0	0	1	10
3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
4	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
5	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
6	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
7	1	1	1	1	0	0	0	0	0	0	1	0	1	1	0	0	1	1	1	0	1	0	1	0	12
8	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
9	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	6
10	1	1	1	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	1	1	0	0	10
11	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
12	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
13	1	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	8
14	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
15	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
16	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
17	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
18	1	1	1	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	1	0	1	1	0	0	9
19	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	6
20	1	1	1	1	0	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	12
21	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
22	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	6
23	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	7
24	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
25	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	1	8
26	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	11
27	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
28	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	7
29	1	1	1	1	1	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	10
30	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
Totals (by picture)	30	30	29	26	4	2	2	0	3	0	6	0	2	2	2	1	6	3	9	2	8	9	4	4	184

Raw Data: 7 year olds

Subject No.																									Totals (by subject)
1	1	1	1	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	1	1	0	0	10
2	1	1	1	1	1	1	0	0	1	1	1	0	1	1	0	0	1	1	1	0	1	1	1	0	18
3	1	1	1	1	0	0	0	0	1	0	0	0	1	0	0	0	1	1	1	0	1	1	1	0	13
4	1	1	1	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	1	1	1	0	11
5	1	1	1	1	0	0	0	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	0	0	10
6	1	1	1	0	1	1	1	0	1	1	0	0	1	1	0	0	1	0	1	0	1	1	1	0	15
7	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	7
8	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	10
9	1	1	1	1	0	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	13
10	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	8
11	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	1	1	1	10
12	1	1	1	1	0	0	0	0	1	0	1	0	1	0	0	0	1	1	0	0	1	1	1	0	12
13	1	1	1	1	0	1	1	0	1	1	0	0	1	1	0	0	1	1	1	0	0	0	0	0	13
14	1	1	1	1	0	0	0	0	1	1	1	0	1	1	0	0	1	1	1	0	0	0	1	0	14
15	1	1	1	1	1	1	0	0	1	1	1	0	1	1	0	0	1	1	0	0	1	1	1	0	17
16	1	1	1	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	14
17	1	1	1	1	1	1	0	0	1	1	1	0	1	1	0	0	1	1	1	0	1	1	1	0	18
18	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
19	1	1	1	1	0	1	0	0	1	1	1	0	1	1	0	0	1	1	1	0	1	1	1	0	17
20	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	0	12
21	1	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	1	1	1	0	0	1	1	0	16
22	0	1	1	1	1	1	1	0	1	1	1	0	1	1	0	0	1	1	1	0	1	1	1	0	18
23	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	0	1	1	1	0	1	1	1	0	19
24	1	1	1	1	0	0	0	0	1	0	1	0	1	0	0	1	1	1	0	0	0	1	0	0	11
25	1	0	1	1	1	1	0	0	1	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	10
26	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	7
27	1	1	1	1	1	0	0	0	1	1	0	0	1	1	0	0	1	1	1	0	1	1	0	0	14
28	1	1	1	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	9
29	1	1	1	1	0	0	0	0	1	1	0	0	1	0	0	0	1	1	1	0	1	1	0	0	12
30	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	19
Totals (by picture)	29	29	29	29	13	16	7	0	21	17	13	0	23	14	12	1	25	21	19	0	22	24	16	1	381

Raw Data: 11 year olds

Subject No.																									Totals (by subject)
1	1	1	1	1	1	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	15
2	1	1	1	0	1	1	0	0	1	0	0	0	1	0	1	0	1	0	0	0	1	1	0	0	11
3	1	1	1	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	14
4	1	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0	1	1	1	0	1	1	1	0	11
5	1	1	1	1	1	1	0	0	1	1	0	0	0	1	1	0	1	1	0	0	1	1	1	0	15
6	1	1	1	1	1	1	0	0	1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	13
7	1	1	1	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	14
8	1	1	1	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	8
9	1	1	1	1	0	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	0	0	0	11
10	1	1	1	1	0	0	0	0	0	0	1	0	1	1	1	0	1	1	1	0	1	1	1	0	14
11	1	1	1	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	1	0	1	1	0	0	15
12	1	1	1	1	0	0	0	0	1	1	0	0	1	0	0	0	1	1	0	0	1	1	0	0	11
13	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	18
14	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	19
15	1	1	1	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	1	0	1	1	1	0	16
16	1	1	1	1	0	1	0	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	17
17	1	1	1	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	1	0	15
18	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	19
19	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	19
20	1	1	1	1	0	0	1	0	1	0	1	0	1	1	1	0	1	1	1	0	1	1	1	0	16
21	1	1	1	0	0	1	0	0	1	1	1	0	1	1	0	0	1	1	1	0	1	1	1	0	15
22	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	18
23	1	1	1	1	0	1	0	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	17
24	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	18
25	1	1	1	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	1	0	1	1	0	0	10
26	1	1	1	1	0	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	12
27	1	1	1	1	1	1	1	0	0	0	0	0	1	0	0	0	1	1	0	0	1	0	0	0	11
28	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	19
29	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	18
30	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	19
Totals (by picture)	30	30	29	24	19	23	12	0	27	21	15	0	29	23	16	0	30	26	19	0	30	26	19	0	448

Reference List

Banet, B., Silverman, C., Allen, M., & Thomson, C. *Levels of representation, part II: experience with representation: the symbol level and the sign level*. Ypsilanti, Michigan: High/Scope Educational Research Foundation, 1972.

Dondis, D. *A Primer of visual literacy*. Cambridge, Mass.: MIT Press, 1973.

Fleming, M. *Instructional illustration--a survey of types occuring in print for four subject areas*. Bloomington, Indiana: Indiana University, 1966.

Friedman, S., & Stevenson, M. Developmental changes in the understanding of implied motion in two-dimensional pictures. *Child Development*, 1975, *46*, 773-778.

Ginsburg, H. & Opper, S. *Piaget's theory of intellectual development*. Englewood, N.J.: Prentice-Hall, 1969.

Gropper, G. Why IS a picture worth a thousand words? *AV Communications Review*, 1963, *11*, 75-95.

Hengen, N. Perception of danger in action illustrations: implications for nonverbal communication education. *Understanding Perception for Better Communication*. Bloomington, Indiana: School of Education, Indiana University, 1970, 1-68

Knowlton, J. On the definition of a picture. *AV Communications Review*, 1966, *14*, (2), 157-183.

Nederveld, P. & Thomson, C. *Levels of representation, part I: experience with real objects: the object and the index level*. Ypsilanti, Michigan: High/Scope Educational Research Foundation, 1972.

Phillips, J., Jr. *The origins of intellect. Piaget's theory* (2nd ed.). San Francisco: W. H. Freeman, 1975.

Rohwer, W., Jr., Ammon, D., & Cramer, P. *Understanding Intellectual Development*. Hinsdale, Illinois: Dryden Press, 1974.

Satterthwaite, L. *An investigation of selected cues for implying motion in non-motion media*. Unpublished doctoral dissertation, Indiana University, 1965.

Wadsworth, B. *Piaget's theory of cognitive development*. New York: David McKay, 1971.

Winer, B. J. *Statistical Principles in Experimental Design*. (2nd ed.). New York: McGraw Hill, 1971.

Effect of Cartoon Illustrations on the Comprehension and Evaluation of Information Presented in the Print and Audio Mode

Edward H. Sewell, Jr.
Department of Performing Arts and Communications
Virginia Polytechnic Institute and State University

Research and Theory Division
Association for Educational Communications & Technology
New Orleans, Louisiana
March 1979

ABSTRACT

The study was designed to investigate the effects of cartoon illustrations on female and male college student comprehension and evaluation of information. Subjects were assigned to one of five treatment groups as follows: (1) printed text, (2) printed text with cartoons, (3) audio-visual presentation, (4) audio only presentation, and (5) visual only presentation. Following the informative presentation, each subject completed a 25-item comprehension test and measures of self-reported enjoyment and helpfulness. The results indicated that there were no significant differences between the printed text, the cartoon text, and the audio-visual presentation in terms of comprehension scores. The cartoon text was perceived as significantly more enjoyable than any other presentation mode. There was a significant interaction between sex and presentation mode with males scoring significantly lower on comprehension in the cartoon text condition.

Effect of Cartoon Illustrations on the Comprehension and Evaluation of Information Presented in the Print and Audio Mode

The use of visual humor in the form of cartoon illustrations is a familiar characteristic in many college textbooks (see for example Hance, Ralph, & Wiksell, 1962; Rein, 1972; and Larson, 1979), and the use of cartoon humor as an instructional technique in the classroom has also been given considerable attention (see Adams, 1974; Brooke, 1973-74; Neie, 1973; and Miner, 1969). The focus of this attention, however, has usually been motivational in its orientation with little evidence advanced to support any positive effects on learning as a result of the use of the cartoon humor. Studies have, in fact, found that the use of humor in most contexts does not result in information gains (Gruner, 1976, pp. 301-304).

In a recent discussion of cartoon humor in magazines, John Peter (1978), a publications consultant and past art director for McCall's, advanced four arguments for the use of cartoons: (1) high readership, (2) enjoyment, (3) increases the visual element, and (4) provides an identity for the publication. In his discussion of these advantages, Peters makes specific reference to the effect of cartoons on reader abilities to remember the content of cartoons, thus suggesting a learning effect.

The present study was designed to investigate the effect of cartoon illustrations on the comprehension and evaluation of information presented in combinations of the print, audio, and visual mode.

Though there has been a significant amount of research investigating the effect of presentation mode on learning, no clear conclusions can be stated based on the research findings (see Hsia, 1971, and Dwyer, 1978 for reviews of the theory and research). One possible reason for the lack of consistent and interpretable results is that such a wide variety of cueing variables are used. A primary function of cueing is to ensure that the intended instructional stimuli are attended in such a way that they can be easily stored in memory and recalled at a later time (see Dwyer, 1978, pp. 152-175 for a review of the theory and research on cueing).

Cartoons serve a cueing function, but little research has centered on the use of cartoons as a means of facilitating learning. Lumsdaine & Gladstone (1958) used cartoons and humorous auditory material in a slide-film technique to teach phonetic alphabet symbols. The introduction of the cartoons and auditory humor resulted in decreased learning.

The use of cartoons in a slide-tape presentation was studied by Baker & Popham (1965) and Popham (1969). Measures of both comprehension and affective reactions were included in the studies, and no comprehension differences were found between presentations with cartoons and those without cartoons. In the second study, affective reactions did favor the cartoon presentation.

Kauffman & Dwyer (1974) studied the use of cartoons and photographs in in-service training. Cartoons were more effective than realistic photographs in facilitating both immediate and delayed retention of information. In addition, a majority of the subjects indicated that they learned more from the cartoon presentation and would prefer to receive cartoon illustrated instruction rather than photographic illustrated instruction.

Using a programmed instructional package with one version containing cartoon supplements, Freisinger (1976) found no differences on either skill performance or affective responses to the subject matter or presentation format.

The effect of the sex of the learner has been almost totally neglected in studies of the relationship between presentation mode and learning. In the one study which used high school or adult subjects and focused on the effect of sex, Dwyer (1971) found no differences in learning from several types of visual presentations.

The present study attempted to investigate (1) the effect on the comprehension and evaluation of information presented in several combinations of print, audio, and visual formats, (2) the effect of the sex of the learner on the comprehension and evaluation of information presented in several combinations of print, audio, and visual formats, and (3) any interactions between presentation mode and sex of the learner.

METHOD

Subjects. Subjects were 150 student volunteers enrolled in several sections of basic communications classes as Virginia Polytechnic Institute and State University. Half the subjects were female, and half were male. While no coursework credit was given for participation in the study, the students were told that if their final grade average in the course was borderline, they would be given the higher grade in the final averaging.

Materials. The materials developed for use in this study were (1) a basic text, (2) an audio recording of the basic text, (3) cartoons based on the basic text, and (4) slides of the cartoons.

A basic text, "The Library at Virginia Tech," was developed from information available in the library as well as with the aid of several members of the library staff. The text described four keys helpful in unlocking the resources of the library.

An audio cassette recording was made of a female reading the basic text.

No attempt was made to provide vocal variety in the recorded version other than that normally associated with a person reading informative material aloud.

A series of forty-three cartoons was created and drawn by a student whose cartoons appeared in each issue of the campus newspaper. The cartoons, which featured characters from one of his regular strips, were created specifically for this study to accompany the content of the basic text.

Black-and-white slides were made from each of the forty-three cartoons.

Treatments. Subjects received one of five experimental treatments. The first treatment was a simple printed copy of the basic text. The booklet looked like a handout one might receive in a class or upon entering the library (see Appendix C for some sample pages). The booklet was seven double-spaced pages, and there were no visuals or headings to provide visual cues about changes in the subject content.

The second treatment was a booklet which combined the basic text with the forty-three cartoons (see Appendix C for some sample pages). This booklet was eighteen pages in length.

The third treatment was an audio-visual presentation which consisted of the audio recording of the text cued to the slides of the cartoons. The total presentation was presented by means of a slide projector connected to a cassette recorder.

The fourth treatment was the audio recording alone, and the fifth treatment was the slide presentation alone.

Dependent Variables. There were three dependent variables in the study. A 25-item multiple-choice comprehension test over the contents of the basic text was developed to measure subject comprehension.

Ten bipolar adjective pairs were used to measure affective responses to the presentation. Factor analysis suggested that there were two well defined factors which were then used as separate affective measures. One factor, called "Enjoyment," contained six adjective pairs (exciting/boring, interesting/dull, humorous/serious, enjoyable/unenjoyable, happy/sad, and unusual/ordinary). The second factor, "Helpfulness," contained three adjective pairs (helpful/unhelpful, informative/uninformative, and worthwhile/worthless).

Procedures. Subjects were randomly assigned to treatments with equal female and male representation in each treatment. The presentations were in the evening with order of presentation being randomly determined.

Subjects were told that they were to read/watch/listen to an informative presentation on the library, and after the presentation they would be tested over the presentation. Time was controlled so that all subjects were given equal time between the beginning of the presentation and the time at which the test was administered. Since those subjects reading the printed texts worked at their own pace, they were instructed to close their booklets when they had read the complete text once.

Statistical Analysis. A 2 x 5 multivariate analysis of variance was performed with two levels of sex and five treatment levels; comprehension scores, self-reported "Enjoyment," and self-reported "Helpfulness" were the dependent variables. Following significant multivariate tests, univariate F ratios were calculated, and following significant univariate tests, multiple comparisons using the Scheffé method were computed.

The .05 level of significance was required for all tests, and the significance level is reported with each separate analysis.

The statistical analyses were performed using programs from both SAS (Barr, *et*. *al*., 1976) and SPSS (Nie, *et*. *al*., 1975).

RESULTS

The reliability of the comprehension test, determined by the KR-20 formula, was found to be 0.78.

The results of the factor analysis discussed earlier is presented in Appendix A. One of the ten bipolar items did not fall into either factor, and was thus not used in any further statistical analysis.

The intercorrelations of the three dependent measures are presented in Table 1. Intercorrelations between Comprehension and Enjoyment, and between Enjoyment and Helpfulness were significant. Comprehension and Helpfulness were not significantly related.

Table 1
Intercorrelations of Dependent Measures

Measure	Comprehension	Enjoyment	Helpfulness
Comprehension	1.000		
Enjoyment	0.276*	1.000	
Helpfulness	0.086	0.338*	1.000

*p<.01

Analysis of Variance. The multivariate analysis of variance was used to determine the main effects of presentation mode and sex on the three dependent measures. Univariate analysis of variance was used when the multivariate analysis was significant, and when the univariate analsysis was significant, *post hoc* analysis of pairwise comparisons was performed using Scheffe.

The raw cell means on the three dependent measures are presented in Appendix B.

The main effect of sex was not significant. The main effect of presentation mode was significant, multivariate $F(12,365)=14.94$, $p<.001$. Univariate F ratios were computed to assess the individual significance of the three dependent measures. All three of the univariate F ratios were significant at $p<.001$: for Comprehension, $F(4,149)=28.27$; for Enjoyment, $F(4,149)=10.88$; for Helpfulness, $F(4,149)=6.31$.

The sex x presentation mode interaction was significant. The multivariate overall effect was $F(12,365)=2.18$, $p<.01$. Only one of the univariate F ratios, for Comprehension, $F(4.149)=4.49$, was significant at $p<.01$.

Post Hoc Analysis. The differences among means for the two dependent variables with significant differences as computed using Scheffe are presented in Tables 2 and 3.

The differences between the printed text, cartoon printed text, and audio-visual presentation were not significant for Comprehension. The printed text and cartoon printed text means were significantly greater than the means for the audio and the visual presentations, and the means for the audio-visual and audio presentations were significantly greater than the mean for the visual presentation.

On the Helpfulness measure, the differences between printed text, cartoon printed text, audio-visual presentation, and audio presentation were not significantly different. Self-reported Helpfulness was significantly lower on the visual presentation than on any other treatment condition.

While there were no significant differences on the Enjoyment measure when using the Scheffe procedure, there were differences when using the Tukey LSD procedure. These differences are reported in Table 4. The cartoon printed text was perceived as significantly more enjoyable than the printed text, the audio presentation, and the visual presentation.

Table 2

Differences Among Means on Comprehension
Scheffé Procedure

		Printed Text	Cartoon Text	Audio-Visual	Audio Only	Visual Only
Printed Text	87.9	---	2.6	7.9	11.9**	24.7**
Cartoon Text	85.3		---	5.3	9.3*	22.1**
Audio-Visual	80.0			---	4.0	16.8**
Audio Only	76.0				---	12.8**
Visual Only	63.2					---

*p<.05
**p<.01

Table 2

Differences Among Means on Helpfulness
Scheffé Procedure

		Audio Only	Cartoon Text	Audio-Visual	Printed Text	Visual Only
Audio Only	11.3	---	0.8	0.8	0.9	2.9**
Cartoon Text	10.5		---	0.0	0.1	2.1*
Audio-Visual	10.5			---	0.1	2.1*
Printed Text	10.4				---	2.0*
Visual Only	8.4					---

*p<.05
**p<.01

Following the significant Sex x Presentation Mode interaction, the *t*-test was used for individual comparisons between the means. The nature of the interaction is ploted in Figure 1. On comprehension, females ($\overline{X}$=91.467) scored significantly higher than males ($\overline{X}$=79.200) in the Cartoon Text treatment. ($\underline{t}$=3.11, df=28, $\underline{p}$<.01). In one other treatment condition, Visual Only, male comprehension scores ($\overline{X}$=66.667) were significantly higher than female comprehension scores ($\overline{X}$=59.733) with $\underline{t}$=-2.09, df=28, $\underline{p}$<.05.

Table 4

Differences Among Means on Enjoyment
Tukey LSD Procedure

		Cartoon Text	Audio-Visual	Printed Text	Visual Only	Audio Only
Cartoon Text	9.3	---	1.3	1.5*	1.8**	4.3**
Audio-Visual	8.0		---	0.2	0.5	3.0
Printed Text	7.8			---	0.3	2.8
Visual Only	7.5				---	2.5
Audio Only	5.0					---

*$\underline{p}$<.05
**$\underline{p}$<.01

DISCUSSION

Of primary importance was the finding that there were no significant differences between the printed text, the cartoon text, and the audio-visual presentation in terms of the comprehension of the content. This result would seem to lend support to studies such as Van Mondfrans & Travers (1964) and Main & Griffiths (1977) which have found that information can be just as easily processed in one of several sensory modalities or combination of modalities.

Additionally, the two printed formats (without and with cartoons) resulted in significantly better comprehension than either the audio or the visual presentation, and the visual presentation resulted in significantly less comprehension than any other presentation mode. The inability of the visual presentation should come as no surprise since it was lacking in the essential information available in the other four presentation modes. It served mainly as a control group.

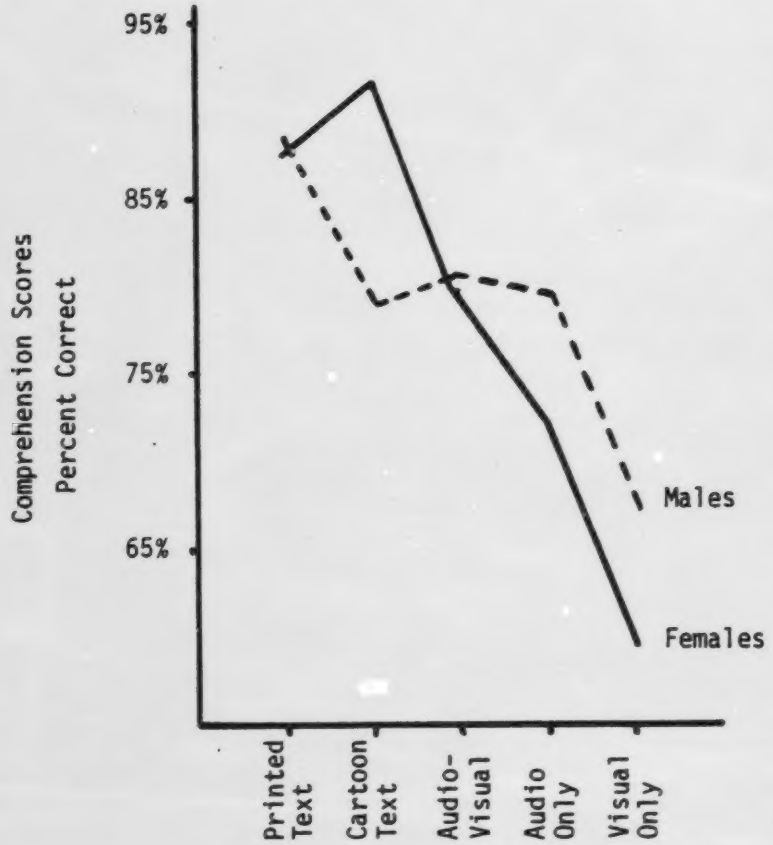


Figure 1
Sex x Presentation Mode Interaction

A lack of any relationship between the perceived helpfulness of a presentation and actual comprehension was demonstrated by the lack of any significant differences on self-reported Helpfulness in any condition but the visual presentation.

The cartoon text was perceived as significantly more enjoyable than any other presentation mode. This result does not support Freisinger's (1976) finding that the addition of a cartoon embellishment resulted in no affective differences. The difference between the two studies could come from a variety of sources. For one thing, the cartoons used in the present study were to some degree "student tested" in that the characters were familiar to the student population being studied since they appeared in a regular cartoon strip in the student newspaper. The nature of the cartoons used in the Freisinger study were not specified.

Of great interest was the significant interaction between Sex and Presentation Mode, accounted for by differences in comprehension scores. To hypothesize that the males were distracted by the cartoons would go against a fairly stable body of research on distractability (see Halley, 1975 for a review of some of this literature). One possible explanation worth further exploration might be that the males were indeed not "distracted" by the cartoons in the printed text, and as a result they missed visual imagery cues which could have been used to recall specific information. This would account for their slightly better performance in the audio-visual condition since the slides of the cartoon were not spatially related to the informative presentation. When only one sensory modality was used, males scored the same as females (printed text) or higher than females (audio and visual only).

In summary and conclusion, the results of the present study would seem to suggest that if comprehension is the only goal of instruction, the less expensive printed text is just as effective as the more expensive illustrated text or audio-visual presentation. If, however, there is a concern for the student to enjoy the presentation while learning the information, there is some support for the value of the cartoon illustrated text. The overall equality of the printed text, cartoon illustrated text, and audio-visual presentation would seem to make final choices largely a matter of personal preference and budget limitations.

REFERENCES

Adams, W. J. The use of sexual humor in teaching human sexuality at the university level. Family Coordinator, 1974, 23, 365-368.

Baker, E. L., & Popham, W. J. Value of pictorial embellishments in a tape-slide instructional program. AV Communication Review, 1965, 13, 397-404.

Barr, A. J., Goodnight, J. H., Sall, J. P., & Helwig, J. T. A User's Guide to SAS 76. Raleigh, N.C.: SAS Institute, Inc., 1976.

Brooke, L. A. Let's keep our basic business students interested. Balance Sheet, 1973/74, 55, 173-174, 188.

Dwyer, F. M. Visual learning: an analysis by sex and grade level. California Journal of Educational Research, 1971, 22, 170-176.

Dwyer, F. M. Strategies for Improving Visual Learning. State College, PA.: Learning Services, 1978.

Freisinger, S. D. The effect of cartoon-embellished programed textual instruction on students' skill learning and affective learning. Paper presented at the Association for Educational Communications and Technology, 1976. ERIC Document ED 128 008.

Gruner, C. R. Wit and humor in mass communication. In T. Chapman & H. Foot (Eds.), Humor and Laughter. New York: John Wiley & Sons, 1976.

Halley, R. D. Distractability of males and females in competing aural message situations: a research note. Human Communication Research, 1975, 2, 79-82.

Hance, K. G., Ralph, D. C., & Wiksell, M. J. Principles of Speaking. Belmont, CA.: Wadsworth Publishing Company, 1962.

Hsia, H. J. The information processing capacity of modality and channel performance. AV Communication Review, 1971, 19, 51-75.

Kauffman, S. P., & Dwyer, F. M. Effectiveness of cartoons and photographs in in-service training. California Journal of Educational Research, 1974, 24, 197-204.

Larson, C. U. Persuasion: Reception and Responsibility (2nd ed.). Belmont, CA.: Wadsworth Publishing Company, 1979.

Lumsdaine, A. A., & Gladstone, A. Overt practice and audiovisual embellishments. In M. A. May & A. A. Lumsdaine (Eds.), Learning from Films. New Haven: Yale University Press, 1958.

Main, R. E., & Griffiths, B. Evaluation of audio and pictorial instructional supplements. AV Communication Review, 1977, 25, 167-179.

Miner, M. E. Charlie Brown goes to school. English Journal, 1969, 59, 1183-1185.

Neie, V. E. The lighter side of science. Science Activities, 1973, 10, 35-37.

Nie, N. H., Hull, C. H., Steinbrenner, K., & Bent, D. H. Statistical Package for the Social Sciences (2nd ed.). New York: McGraw-Hill Book Company, 1975.

Peter, J. A sense of humor. Folio, March 1978, pp. 41-54.

Popham, W. J. Pictorial embellishments in a tape-slide instructional program. AV Communication Review, 1969, 17, 28-35.

Rein, I. J. Rudy's Red Wagon: Communication Strategies in Contemporary Society. Glenview, IL.: Scott, Foresman and Company, 1972.

Van Mondfrans, A. P., & Travers, R. M. W. Learning of redundant material presented through two sensory modalities. Perceptual and Motor Skills, 1964, 19, 743-751.

APPENDIX A

Results of Factor Analysis

	Factor I	Factor II	h^2
exciting/boring	.842*	.156	.73
interesting/dull	.853*	.173	.76
humorous/serious	.689*	-.301	.57
enjoyable/unenjoyable	.856*	.147	.75
happy/sad	.697*	.112	.50
unusual/ordinary	.768*	-.112	.60
helpful/unhelpful	.045	.907*	.83
informative/uninformative	-.045	.871*	.76
worthwhile/worthless	.233	.762*	.64
involving/uninvolving	.591	.304	.44
Proportional contribution to common variances	4.128	2.447	

The analysis was a principal components with varimax rotation. Each factor was required to have at least two items loaded at .600 or above with no secondary loadings at .400 or above.

APPENDIX B

Cell Means

		Female	Male	Row Means
Printed Text	Comprehension	87.467	88.267	87.867
	Enjoyment	7.267	8.400	7.833
	Helpfulness	10.600	10.134	10.367
Printed Cartoon Text	Comprehension	91.467	79.200	85.333
	Enjoyment	9.200	9.467	9.333
	Helpfulness	10.867	10.134	10.500
Audio-Visual Presentation	Comprehension	79.467	80.533	80.000
	Enjoyment	8.000	8.000	8.000
	Helpfulness	10.333	10.667	10.500
Audio Only Presentation	Comprehension	72.000	79.200	76.000
	Enjoyment	5.333	4.600	4.967
	Helpfulness	11.800	10.733	11.267
Visual Only Presentation	Comprehension	59.733	66.667	63.200
	Enjoyment	6.667	8.333	7.500
	Helpfulness	7.733	9.000	8.367
Column Means	Comprehension	78.187	78.773	
	Enjoyment	7.293	7.760	
	Helpfulness	10.267	10.133	

APPENDIX C

Samples of the Printed Texts

THE LIBRARY AT VIRGINIA TECH

The Newman Library at Virginia Tech is there to serve the needs of the students and faculty by providing the tools for the basic instruction and research needs of the university community.

To help you more effectively and efficiently use the resources of the library, we would like to introduce you to several valuable keys which can help unlock the mysteries of the library. There are four major keys we will talk about.

First is the key to knowing where to locate the books and periodicals you will need.

Second is the key known as the card catalog system.

Third is the key known as the library classification systems.

Finally, there is the key to some rules and services which help to make the library a better place for you and everyone else.

Let's begin with that first key--location of books and periodicals.

You should know that in addition to the main Newman Library, there are two branch libraries with books and other materials in specialized subject areas. The books and resource materials in branch libraries are for the use of all students.

The Architecture Library is located in Cowgill Hall and contains materials about architecture and visual arts. All the books in the Architecture Library are catalogued in the main library, but after the usual call number, you will find the abbreviation ARCH.

You will find the Geology Library in Derring Hall. Not everything in this branch library is written on rock. All the books and learning resources located in the Geology Library are also catalogued in the main card catalog of the Newman Library, but after the call number you will find the abbreviation GEOL.

Most of the books, of course, are located in the main Newman Library, and we will focus most of our attention on finding books and periodicals there.

In the main library there are three subject divisions: Humanities, Social Science, and Science and Technology.

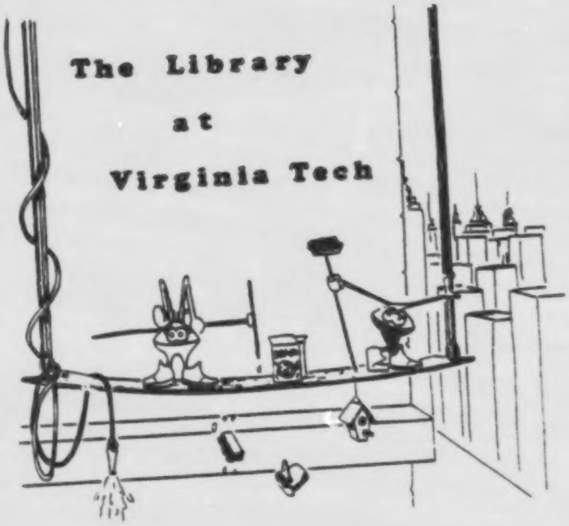
Humanities books and periodicals are located on the 1st and 2nd floors of the Newman Library. Here you will find books and periodicals which provide you with information about philosophy and religion, history, music, language and literature, theatre, and communications.

Books about Van Gogh, Matisse, Bach, Beethoven, McLuhan, and Kierkegaard will be found in the Humanities.

Did you know that the library receives newspapers from around the world as well as from across the U.S. and the Commonwealth of Virginia? These newspapers are available in the Humanities Division of the 2nd floor.

Books and periodicals in Science and Technology are located on the 3rd floor of Newman Library. Here you will find information about the natural sciences, medicine, agriculture, technology, mathematics, and military science. This is the place to come if you need a solution to that impossible equation, or have you ever thought of building your own bridge?

The Social Science Division is located on the 4th floor and contains books and periodicals about psychology, law, geography, anthropology, political science, sociology, business, and education.


The Library
at
Virginia Tech

The Newman Library at Virginia Tech is there to serve the needs of the students and faculty by providing the tools for the basic instruction and research needs of the university community.

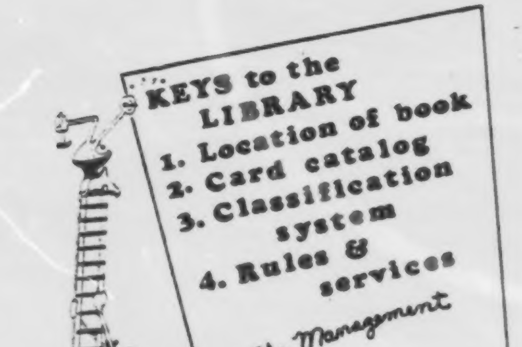
To help you more effectively and efficiently use the resources of the library, we would like to introduce you to several valuable keys which can help unlock the mysteries of the library. There are four major keys we will talk about.

First is the key to knowing where to locate the books and periodicals you will need.

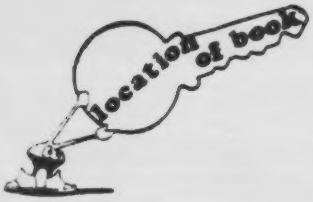
Second is the key known as the card catalog system.

Third is the key known as the library classification systems.

Finally, there is the key to some rules and services which help to make the library a better place for you and everyone else.


KEYS to the
LIBRARY
1. Location of book
2. Card catalog
3. Classification
system
4. Rules &
services
The Management

2



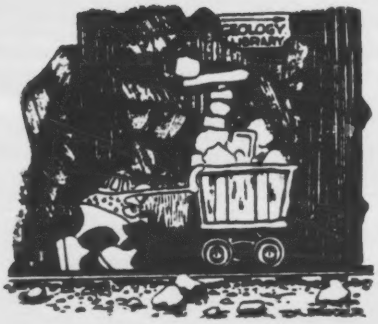
Let's begin with that first key--the location of books and periodicals.

You should know that in addition to the main Newman Library, there are two branch libraries with books and other materials in specialized subject areas. The books and resource materials in branch libraries are for the use of all students.

ARCHITECTURE LIBRARY

The Architecture Library is located in Cowgill Hall and contains materials about architecture and visual arts. All the books in the Architecture Library are catalogued in the main library, but after the usual call number, you will find the abbreviation ARCH.

3



You will find the Geology Library in Derring Hall. Not everything in this branch library is written on rock. All the books and learning resources located in the Geology Library are also catalogued in the main card catalog of the Newman Library, but after the call number you will find the abbreviation GEOL.

Most of the books, of course, are located in the main Newman Library, and we will focus most of our attention on finding books and periodicals there.

In the main library there are three subject divisions: Humanities, Social Science, and Science and Technology.

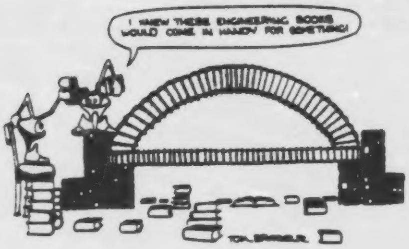
Humanities books and periodicals are located on the 1st and 2nd floors of the Newman Library. Here you will find books and periodicals which provide you with information about philosophy and religion, history, music, languages and literature, theatre, and communications.

Books about Van Gogh, Matisse, Bach, Beethoven, McLuhan, and Kierkegaard will be found in the Humanities.

TOM BRODEUR

Did you know that the library receives newspapers from around the world as well as from across the U.S. and the Commonwealth of Virginia? These newspapers are available in the Humanities Division on the 2nd floor.

Books and periodicals in Science and Technology are located on the 3rd floor of Newman Library. Here you will find information about the natural sciences, medicine, agriculture, technology, mathematics, and military science. This is the place to come if you need a solution to that impossible equation, or have you ever thought of building your own bridge?


I KNEW THESE ENGINEERING BOOKS WOULD COME IN HANDY FOR SOMETHING!
TOM BRODEUR

I DON'T THINK THIS IS WHAT "SOCIAL SCIENCE" MEANS!
TOM BRODEUR

The Social Science Division is located on the 4th floor and contains books and periodicals about psychology, law, geography, anthropology, political science, sociology, business, and education.

"THEN INSERT FILM INTO TAKE-UP REEL..."
HOW TO USE
TOM BRODEUR

Several other types of special materials are located on the 4th floor along with the Social Science Division.

You will find most of the microforms (such as microfilm, microfiche, and microcards) located there.

WE'LL START OUR INVASION HERE!

There is also a map library.

PART IV:

MEASUREMENT AND ASSESSMENT TECHNIQUES

IR007107

TITLE: Measuring Attitudes and Instructional Development: Why and How

AUTHOR: Dr. Michael Simonson
Associate Professor of Secondary Education
321 Curtiss Hall
Iowa State University
Ames, Iowa 50011
(515) 294-6840

Media specialists and instructional developers generally ask three questions when the measurement of attitudes is promoted as a component of the instructional design process.

- Why are attitudes important in the teaching/learning process?
- Why is it important to measure attitudes?
- How are attitudes measured?

This paper will attempt to answer these questions.

Part I. Importance of Attitudes and Attitude Measurement

When the instructional developer designs a classroom activity there should be at least two categories of learning outcomes in mind--those directed toward cognitive goals, and those related to the attitudes of the learner. There is little need to discuss the rationale for the importance of information acquisition by a learner as a result of instruction. The need for establishing attitudinal goals and for planning activities designed to produce affective outcomes in learners as a consequence of an instructional sequence is a little more difficult to explain to many. However, it has become increasingly apparent to many involved in educational technology research that one of the major, unique consequences of mediated instruction is not directed toward knowledge gain. Rather, instruction from television, film and slides often produces certain attitudinal positions in students not necessarily found when media are not used in teaching. In a recent review of educational technology research it was found that when attitude hypotheses were tested, over fifty percent of the time desired attitudinal positions or changes were produced (Simonson, 1977; 1979a; 1979b). In other words, in about one half of the experiments reviewed, the attitude toward instructional method or content area possessed by students were either favorable, or changed in a desired direction, after instruction that was delivered primarily by media.

Still, the instructional developer may say, "So what, the purpose of instruction is to promote the acquisition of knowledge and skills, and it doesn't make any difference if students like or dislike what they learn, or how they are taught."

This persuasive and prevalent, if narrow, argument is somewhat difficult to refute. The most powerful rationale for the need to promote attitude positions in learners would be to demonstrate a direct relationship between attitudes and achievement, or liking and learning. And as a matter of fact, numerous researchers have identified just such a relationship (Simonson and Bullard, 1978; Simonson, 1977; Levy, 1973; Fenneman, 1973; Perry and Kopperman, 1973; and Greenwald, 1966, 1965, for example). However, most educational researchers are very reluctant to claim that there is any cause and effect linkage between these two learner variables.

Probably the development of a positive attitude in learners should be a desirable end in itself. There are several reasons why attitudes of learners should be important. First, most instructional developers would agree that there are cases when it is legitimate, and important, to urge learners to accept the truth of certain ideas. In other words, to promote an attitudinal position. Second, while the strength of the relationship between attitudes and achievement is unclear, it makes common sense that students are more likely to remember information, seek new ideas, and continue studying, when they react favorably to an instructional activity, or "like" a certain content area. Third, there are some instances when influencing the attitudes of students is not desirable, so instructional developers should be aware of what techniques do affect attitudes. In this way possible bias can be recognized and eliminated. Last, attitudes toward instruction felt by learners can tell the developer a great deal about the impact of that instruction on the learning process. In other words, we need to assess the opinions of our students toward the learning activities we are subjecting them to, if for no other reason than to improve the quality of our procedures. Whatever the reason, attitudinal outcomes should be important considerations for the designer of instruction.

DEFINING ATTITUDES: Before attitudes can be measured, we must identify what this term means, and what contributes to additional positions in learners. Attitude has been a difficult concept to adequately define, primarily because it has been defined by so many, but also because of its many lay uses and connotations. One of the earliest definitions of attitude was proposed by Thomas and Znaniecki (1918). They defined attitude as:

> A mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related (Thomas and Znaniecki, 1918).

In other words, while attitudes are latent and not directly observable in themselves, they do act to organize, or to provide direction to, actions and behaviors that are observable. Also, attitudes vary in direction, either positive or negative; in degree, the amount of positiveness or negativeness; and in intensity, the amount of commitment with which a position is held (Fleming and Levie, 1978).

Additionally, attitudes have three components: affective, cognitive, and behavioral (Zimbardo and Ebbeson, 1970). The affective component is said to consist of a person's evaluation of, liking of, or emotional response to some object or person. The cognitive component is conceptualized as a person's beliefs about, or factual knowledge of, the object or person. The behavioral component involves the person's overt behavior directed toward the object or person.

Part II. How Attitudes are Measured

Since attitudes are defined as latent, and not observable in themselves, the instructional developer must identify some behavior that would seem to be representative of the attitude in question, so that this behavior might be measured. This characteristic of attitude measurement is justifiably the most criticized limitation of this area of educational evaluation. However, without going into the question of the over-all validity of attitude measurement, there are several generally recognized procedures used to determine an individual's or group's attitude toward some object or person. It is those procedures that are outlined below.

CHARACTERISTICS OF MEASUREMENT: Before procedures for measuring attitudes are discussed, there are several characteristics of measurement, in general, that should be considered in order to determine if an evaluation technique is an effective one. Basically, attitude measures should:

- be valid -- In other words, is the instrument appropriate for what needs to be measured?
- be reliable -- Does the measure yield consistent results?

- be fairly simple to administer, explain, and understand -- Generally, the measures that yield a single "score" of an attitude position epitomize the intent of this characteristic, although the single "score" may be deficient in meeting the intent of other characteristics of good measurement.
- be replicable -- In other words, someone else should be able to use a measure on a different population, or in a different situation, to measure the same attitude.

CATEGORIES OF ATTITUDE MEASUREMENT: Basically, there are four categories, or approaches, for collecting attitude information. These approaches are:

- self-reports, where the members of a group report directly about their own attitudes,
- reports of others, where others report about the attitudes of a person or group,
- sociometric procedures, where members of a group report about their attitudes toward one another, and
- records, which are systematic accounts of regular occurrences, such as attendance reports, sign-in sheets, library check-out records, and inventories.

Within each of these categories there are one, or more, strategies for measuring attitude-related behaviors. Most commonly, attitude measurement is accomplished by one of the following techniques.

- questionnaires
- rating scales
- interviews
- written reports
- observations
- sociometrics

PROCESS FOR ATTITUDE MEASUREMENT: Any attempt at measurement, including the evaluation of attitude, requires that a systematic process be followed. Such procedures will not guarantee an effective measurement, but they do increase the likelihood of this occurring, considerably. Generally, there are six steps to be followed during the attitude measurement process.

1. Identify Construct to be Measured - A learner could conceivably have an attitude position toward any object, situation, or person. When instruction is designed, those attitudes that are important to the learning activity should be clearly identified and defined. For example, if the developer of instruction wanted to ascertain the impact of a certain type of media on learners, the construct "attitude toward instruction by film" could have been an attitude outcome that was evaluated.

2. Find an Existing Measure of the Construct - Once a certain attitude has been identified, the designer should attempt to locate an instrument that will measure the relevant construct. Generally, such tests will have been tried out in other instructional situations and should include some statement of reliability and validity. Additionally, instructions for administration of the test selected should be included. This will simplify the job of using the instrument for the instructional developer.

The most obvious disadvantage to using a pre-prepared measure is that it may not be evaluating the specific attitude construct being studied. Even if this is the case, it may sometimes be possible to extract valuable information from an instrument designed to test an attitude position similar to the one of specific interest.

There are a large number of sources for finding existing attitude instruments. Buros' Mental Measurements Yearbook is probably the most widely known. Of equal value is Buros' Tests in Print. Other reference sources for attitude tests include the CSE Test Evaluation Series, Robinson's Measures of Psychological Attitudes, Rosen's Attitudes Toward School and School Adjustment, and Knapp's Omnibus of Measures Related to School Based Attitudes (see reference list for citations).

3. Construct an Attitude Measure - Obviously, if no existing measure of the relevant attitude is available the instructional developer will need to construct his/her own test. This is probably the most difficult and time consuming step in the attitude measurement process. Often it is the most critical. The test developer should apply a large degree of patience and effort during these procedures.

Of the many types of attitude measurement possible (see above), one technique widely used that seems to possess most of the characteristics of a good measure is the "Agreement, or Likert-type, Scale." This technique uses statements about the attitude object that are either clearly favorable or unfavorable. Each subject responds to each test item according to his/her perceived attitude "intensity" toward the statement. Often, students are asked to answer each test item on a five point scale that has responses that vary in the degree of agreement to the statement from strongly disagree to strongly agree (see Figure 1). One advantage of this technique is the ease of scoring and summarizing the information obtained. (For a complete listing of the steps necessary to construct an agreement scale see Table 1; for a more complete description of how various types of attitude measures are designed see Henerson, 1978.)

When a test is constructed locally it is critical that reliability and validity information be collected for the measure. Of these two concepts, validity (ie. appropriateness of instrument) is the most difficult to determine. Validity for a test depends on a number of factors, such as the type of test and its intended use. Basically, there are four categories of validity.

- Construct validity (or the extent to which you can be sure a measure represents the attitude construct whose name appears in its title) can be determined by:
    1. opinions of judges,
    2. correlations to other measures of the same construct,
    3. measures of criterion group subjects (those who obviously possess the construct), or
    4. appeals to logic.
- Content validity (or the representativeness of the sample of questions included in the instrument) is usually determined by careful analysis of the items in the test. There is no simple process to determine content validity other than a close, thoughtful examination of each item separately, and all items together.
- Concurrent validity (or the agreement of a test with a parallel form of the test on the same topic that was administered at approximately the same time) is determined by correlating the results of two parallel measures of the attitude. This correlation coefficient is reported as an index of concurrent validity.
- Predictive validity (or how well a measure will predict some future behavior) is determined by comparing results of an attitude test to some measure of behavior given at some point in the future. Again, this type of validity is usually expressed by a correlation coefficient found by comparing results of the two measures.

Obviously, determining validity is not a simple task. However, every instructional developer who constructs a test of any type should be acutely aware of the need to develop valid instruments. Since there is no single established method for determining validity, the test originator should use care in constructing, administering and interpreting tests, and their results. Finally, these precautions should be described to, and for, the consumer of the test data.

Reliability, or the ability of a measure to produce consistent results, is usually less difficult to determine than validity. There are several methods of determining reliability that can be easily used by the attitude test developer. The "Test-Retest" method involves the re-administration of the instrument to the target group and correlating the results. The "Split-Half" method uses an arbritrary division of the instrument into two halves. Results from each half are correlated and reported as a reliability coefficient. "Alternate-Form" reliability involves the correlation of the results of two parallel forms of tests of the same attitude construct. Each subject takes each form and the resulting correlation is reported as a reliability estimate.

Each of these techniques will yield a score from 0.00 to 1.00. The higher the number, the more reliable the test. Generally speaking, reliability coefficients above .70 are considered respectable. Scores above .90 are not uncommon for standardized attitude tests. As with validity, the results of reliability estimation should be reported to the consumer of your testing activities. (For more information on validity and reliability estimation see Anastasi, 1968; Cronbach, 1970; Fitz-Gibbon, 1978; Henerson, 1978; or Talmage, 1976).

4. Conduct a Pilot Study - While it is possible to obtain validity and

reliability data during the actual testing portion of the instructional activity, it is much more logical to at least try out attitude instruments before they are formally used. This should be done in order to obtain appropriate data, but also to uncover minor, but troublesome, administrative problems, such as misspellings, poor wording, or confusing directions.

5. Revise Tests for Actual Use - Results of pilot testing should be used to revise, or "polish," attitude instruments. Once the "bugs" have been eliminated the measure is ready to be used with its intended target audience.

6. Summarize, Analyze, and Display Results - When testing is completed resulting data must be interpreted. Generally, attitude test results should be handled similarly to any other test information. Obviously, numerical data is easier to manipulate than verbal information. Whatever attitude responses have been collected, it is important to summarize, analyze and display the results in such a manner that they are easily and quickly understood by others. For example, raw data should be collected and recorded on some type of summary sheet. A "quick-tally" sheet is often used when data is to be hand scored (see Figure 2). When data is to be machine scored it should be recorded on the familiar 80 column coding sheet. This sheet corresponds to the standard 80 column computer card. An even simpler technique for coding data is to have students respond to attitude test items on a "mark-sense" (optical scoring) sheet. When this type of scoring sheet is used, the raw data can be easily accessed for computerized statistical analysis without the need for intermediate coding steps (see Henerson, 1978).

After data has been collected and coded it should be analyzed. Most researchers consider "Agreement Scale" data to be "Ordinal-Scale" (Ferguson, 1971), so it can by analyzed using standard tests of description and inference. However, data about instructional mode or content area is often useful if it is merely averaged and compared to other averages. In other words, did the class change in average "Attitude Toward India" after viewing the film? Did they react favorably to "Film as a Method of Instruction as Compared to Lecture?" Often merely "eyeballing" average scores or "change scores" will be helpful in analyzing data and in interpreting the impact of mediated instruction on attitudes.

Displaying data is another effective method of analyzing it for the consumer. Charts, graphs, & bar diagrams are examples of data display techniques that are useful in assisting in the understanding of what test results indicate. Whatever the process, the developer of an attitude test should make every effort to decipher the results of the measure, and to explain those conclusions and implications to the consumer of the data. Sometimes this may mean that the developer will have to apply a considerable amount of subjective interpretation to data. This is not necessarily bad. Naturally, the more that opinions can be based on, or replaced by, facts, the more powerful conclusions will be. (See Figure 3 for examples of data display techniques.)

Conclusion

Obviously, attitude measurement is only one of many evaluation techniques for the instructional developer to consider for use. However, since attitudes are "predispositions to respond" they would seem to be related, in some manner and degree, to what happens to, and in, a learner during the instructional process. Admittedly, there are many possible deficiencies in the measurement techniques described. Attitude measurement is certainly not as refined as anyone would like. That does not mean that understanding the impact this construct has on the learning process is not important. In order to facilitate a more complete understanding of teaching and learning the instructional developer should collect as much relevant data about that process as is possible. Attitudes and attitude measurement are a critical component of that understanding.

REFERENCES

Anastasi, A. *Psychological testing*. New York: Macmillan, 1968.

Buros, O.K. (Ed.) *The mental measurements yearbook*. Highland Park, New Jersey: Gryphon Press.

Buros, O.K. (Ed.) *Tests in print*. Highland Park, New Jersey: Gryphon Press.

Cronbach, L. J. Test validation. In R. L. Thorndike (Ed.), *Educational measurement*. Washington, D.C.: American Council on Education, 1970.

Fenneman, G. C. *The validity of previous experience, attitude, and attitude toward mathematics as predictors of achievement in freshman mathematics at Wartburg College*. Unpublished doctoral dissertation, University of Northern Colorado, 1973.

Ferguson, G. A. *Statistical analysis in psychology and education*. New York: McGraw-Hill, 1971.

Fitz-Gibbon, C. T. & Morris, L. L. *How to calculate statistics*. Beverly Hills: Sage, 1978.

Fleming, M. & Levie, W. H. *Instructional message design*. Englewood Cliffs, New Jersey: Educational Technology Publications, 1978.

Greenwald, A. G. Behavior change following a persuasive communication. *Journal of Personality*, 1965, *33*, 370-391.

Greenwald, A. G. Effects of prior commitment on behavior change after a persuasive communication. *Public Opinion Quarterly*, 1966, *29*, 595-601.

Henerson, M. E. Morris, L. L. and Fitz-Gibbons, C. T. *How to measure attitudes*. Beverly Hills, California: Sage, 1978.

Hoepfner, R., et.al. *The CSE test evaluation series*. Los Angeles: Center for the Study of Education, UCLA Graduate School of Education.

Knapp, J. *An omnibus of measures related to school based attitudes*. Princeton, N. J.: Educational Testing Service, Center for Statewide Educational Assessment, 1972.

Levy, J. *Factors related to attitudes and student achievement under a high school foreign language contingency contract*. Unpublished doctoral dissertation. University of Southern California, 1973.

Perry, G. A. & Kopperman, N. *A better chance-evaluation of student attitudes and academic performance, 1964-1972*. Boston: A Better Chance, Inc., 1973.

Robinson, J. P. & Shaver, P. R. *Measures of social psychological attitudes*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan, 1973.

Rosen, P. (Ed.) *Attitudes toward school and school adjustment, grades 4-6*. Princeton, N.J.: Educational Testing Service, 1973.

Simonson, M. R. Attitude change and achievement: Dissonance theory in education. *Journal of Educational Research*, 1977, *70*(3), 163-169.

Simonson, M. R. Media and attitudes: An annotated bibliography of selected research--Part II. *Educational Communication and Technology Journal*, 1979, in press. (b)

Simonson, M. R. & Bullard, J. Influence of student expectations and student sex on predicting academic success. Paper presented at the meeting of the Midwest Area Educational Research Association, Chicago, 1978. *Resources in Education*, ED 114049.

Simonson, M. R. & Burch, G. *Media influence on affective learning*. Paper presented at the meeting of Association for Educational Communication and Technology, Miami Beach, Florida, April, 1977.

Simonson, M. R., Thies, P., & Burch, G. Media and attitudes: Pertinent research published in AV Communication Review. *Educational Communication and Technology Journal*, 1979, in press. (a)

Talmage, H. *Statistics as a tool for educational practitioners*. Berkeley: McCutchan, 1978.

Thomas, W. I. & Znaniecki, F. *The Polish peasant in Europe and America*. Boston: Badger, 1918.

Zimbardo, P. & Ebbesen, E. *Influencing attitudes and changing behavior*. Reading, Massachusetts: Addison-Wesley, 1970.

FIGURE 1 - AGREEMENT SCALE

======	======	======	======	======
1	2	3	4	5
SA	A	U	D	SD
Strongly Agree	Agree	Uncertain	Disagree	Strongly Disagree

FIGURE 2 - "QUICK-TALLY" SHEET

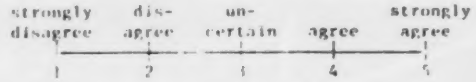
ITEM #	YES	NO	?
1	~~////~~ ///	///	/
2	//	~~////~~ /	////

FIGURE 3: DISPLAYING GROUP DATA

Example. Of the 50 children interviewed, 19 boys and 13 girls reported having taken part in the after school recreation program. These 32 children reported having engaged in the following activities.

	boys	girls	total
handball	19	7	26
bars and rings	16	12	28
team games (baseball, kickball)	17	10	27
handicrafts	13	12	25
chess	8	8	16
checkers	10	8	18

Example. On the following scale, responses were distributed as indicated by the numbers below the scale.



Scale	1	2	3	4	5
Number of respondents	20	6	2	2	24

Were you to calculate the average response, you would obtain 3.07, close to "uncertain." Since 20 people checked "strongly disagree" and 24 checked "strongly agree," it would certainly be misleading to report only that the average response was 3.07, "uncertain"! The *bimodal* character of the responses would have to be reported.

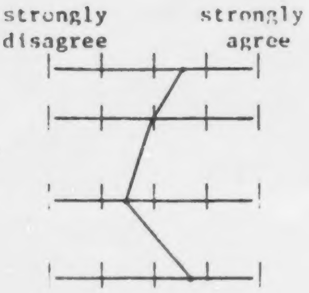
Example. Of the 500 questionnaires mailed to the parents, 225 were returned. The following presents the numbers and percentages of responses item by item.

		Yes	Uncertain	No
1.	Have you been given an explanation of why most classes at school consist of more than one grade level?	137 (61%)	26 (12%)	62 (27%)
2.	Are you in favor of classroom groupings that consist of children of more than one grade level?	170 (75%)	15 (7%)	40 (18%)
3.	Are you in favor of having parent volunteers in your child's classroom?	220 (98%)	3 (1%)	2 (1%)

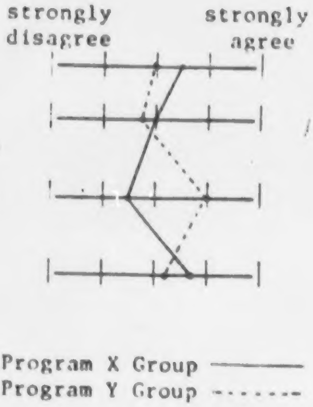
FIGURE 4: DISPLAYING ITEM-BY-ITEM RESULTS

If you have used a questionnaire, you might wish to display the results *on the questionnaire itself.* Given sufficient space on a blank questionnaire, you could record the number of percentage of each response option:

	Yes	No
Did you attend the open house?	☐ 200 (67%)	☐ 100 (33%)
Did you attend the back-to-school night?	☐ 151 (50%)	☐ 149 (50%)



1. I like school a lot.
2. I like school better this year than last year.
3. I think my teacher likes me.
4. I get along with my classmates.



1. I like school a lot.
2. I like school better this year than last year.
3. I think my teacher likes me.
4. I get along with my classmates.

TABLE 1: STEPS FOR DEVELOPING AN AGREEMENT SCALE (Henerson, 1978)

The Agreement Scale

The more common agreement scale also consists of a series of attitude statements. Unlike the ordered scale, however, these sentences do not represent gradations of the attitude. They embody extreme statements, either clearly favorable or clearly unfavorable. The agreement scale achieves a wide range of scores by having respondents report the *intensity* of an attitude. This is accomplished by providing gradations within the response alternatives. The respondents are asked to indicate their agreement with each statement on a 5-point scale:

SA	A	U	D	SD
strongly agree	agree	undecided	disagree	strongly disagree

This question format is a familiar one that can be found in a variety of measures, not just attitude rating scales. It is a popular multiple-choice format that is frequently used in the construction of many types of attitude questionnaires.

Steps for Constructing and Using an Agreement Scale

1. Accumulate a large number of clearly favorable or clearly unfavorable statements about the attitude you wish to measure (approximately 60). As with the statements for the ordered scale, a good source might be prospective respondents.
2. Ask a *pilot* group (50 or more) to respond to these statements. The pilot group should consist of people who are (a) similar to the people whose attitudes you wish to measure, and (b) likely to express the whole range of attitudes you wish the instrument to detect.
3. Score responses by assigning them from one to five points five for most favorable, one for least favorable. This, of course, means responses will be scored differently depending on whether the statement reflects a negative or positive attitude.

Example. Scoring key for two items measuring attitude toward school

School is a waste of time.

SA	A	U	D	SD
1	2	3	4	5

School teaches you things that help in getting a job.

SA	A	U	D	SD
5	4	3	2	1

Note that the highest rating (5) for favorable attitude toward school may be given to strongly disagree (SD) or to strongly agree (SA) depending on how the item is worded.

4. Compute a score for each respondent by totaling the points corresponding to his or her responses.
5. Identify high scorers (top 25%) and low scorers (lowest 25%).
6. Analyze each statement according to how high and low scorers responded to it. The method for accomplishing this step, called "item analysis," is discussed below.
7. Retain those items (approximately 20) which provided good discrimination between high and low scorers.
8. Construct the questionnaire by listing the retained statements in random order.
9. Administer the instrument.
10. Compute a score for each respondent by totaling the scores corresponding to his or her responses.

TABLE 2: STEPS FOR DEVELOPING AN ORDERED SCALE (Henerson, 1978)

The Ordered Scale

The ordered scale consists of a collection of statements that express a range of opinions about an attitude object:

- School teaches you things that help in getting a job.
- Most teachers deeply care about their students.
- Sometimes school can be interesting.
- School is really a bore.
- There are too many rules at school.
- School is the most valuable way I can spend my time.

Before inclusion on the measure, these statements are assigned a "scale value" along an attitude continuum by a panel of judges. An instrument is then constructed which presents the statements in random order without indicating their scale values. The respondents are asked to check those statements with which they agree. Only *opinion statements* are used. *Factual statements* are not included in this kind of measure, since people with differing attitudes could agree to a statement of fact. Note that the respondents are asked to select only those items *with which they agree* and to reject all others.

Steps For Constructing and Using an Ordered Scale

1. Accumulate a large number of statements about the attitude object (approximately 100). A good source for these statements might be interviews of prospective respondents. Make sure that the statements represent a wide range of opinion, including moderate ones.
2. Place each statement on a separate piece of paper.
3. Select a group of judges (preferably 30 or more people like the prospective respondents) and ask each one to sort the statements into 11 piles ranging from highly unfavorable (1) through neutral (6) to highly favorable (11). *Make sure the judges understand that they are classifying the statements and not indicating their agreement or disagreement with them.*
4. Throw out statements that have been placed in widely differing piles. These are ambiguous statements.
5. For each remaining statement, arrive at a scale value by computing the median or mean position of the assignments by the judges.
6. Select a set of statements (under 25) whose scale values will give you a spread that evenly covers the continuum from highly unfavorable to highly favorable.
7. Construct the questionnaire by listing the statements in *random order*. Do not indicate scale values on the instrument.
8. Administer the instrument, instructing respondents to indicate with which statements they agree.
9. Compute a score for each respondent by finding the mean of the scale values of the statements selected by that respondent.

TABLE 3: STEPS FOR DEVELOPING A SEMANTIC DIFFERENTIAL (Henerson, 1978)

Steps for Constructing and Using a Semantic Differential

1. Determine the attitude object(s) you wish to investigate.
2. Select appropriate adjective pairs (approximately 10). You may wish to select from the list provided at the end of this chapter or from DiVesta's list if it suits your students. You may, on the other hand, wish to make up your own list.
3. Write the attitude-object word or phrase at the top of the page and place the adjectives beneath it. If you are examining more than one attitude object, use the same adjective order for each attitude object, and keep the words in the same position. Provide "random polarity." This means that the adjective pairs should *not* be listed so that all positive responses fall on one side and all negative responses on the other.

Example

How do you feel about each of these subjects? Place an X on one of the seven lines between each pair.

physical education

bad								good
friendly								unfriendly
fair								unfair
sour								sweet

art

bad								good
friendly								unfriendly
fair								unfair
sour								sweet

---

4. Instruct the respondents about how and where to mark their ratings. They should be instructed to respond quickly and on the basis of their first impressions. You may find that some people are not comfortable responding to a concept (e.g., physical education, art) with seemingly inappropriate adjectives (e.g., sweet-sour). You will have to reassure such people that this type of scale calls for *impressions*, not studied responses. If you are using the instrument with children, you would do well to give them some practice with other concepts (attitude objects) before they make their ratings on the concepts you want to measure.
5. Compute a person's score by assigning a "1" to responses indicating the most negative response, a "7" to the most positive response, and scoring intermediate responses from "2" to "6" accordingly. A person's score for any one attitude object is the average of responses to the attitude pairs.

Here are more adjective pairs you might use for constructing semantic differential rating scales:

negative	positive	negative	positive
boring	interesting	unhealthy	healthy
uninformative	informative	dull	lively
confusing	clear	weak	strong
irrelevant	relevant	unfair	fair
superficial	profound	dirty	clean
biased	objective	worthless	valuable
purposeless	purposeful	useless	useful
closed	open	passive	active
tense	relaxed	static	dynamic
unhappy	happy	unfriendly	friendly
angry	calm	wrong	right
cold	warm		

NEEDS ASSESSMENT

[illegible]

Dean R. Spitzer
Assistant Director
South Central Regional Medical Education Center
VA Medical Center
St. Louis, Missouri 63125

Abstract

Needs assessment is a fundamental part of evaluation, and yet it has not developed as rapidly in sophistication as other evaluation methodologies. This is largely due to the difficulty in measuring real needs, as opposed to wants, interests, etc. This paper discusses some of the critical issues in needs assessment, and outlines some of the fundamental questions that must be answered in order to design a valid and reliable needs assessment study. Emphasis is placed on the need for considering resources and constraints in designing a realistic approach to needs assessment.

NEEDS ASSESSMENT

Needs assessment is a large class of activities that are aimed at determining the needs that exist among a certain group of people. Sometimes these needs are obvious and readily observable; othertimes, these needs are hidden and not perceivable without fine instrumentation. If we assume, as I think we must, that all programs exist to serve people with needs, then the importance of needs assessment can be seen. In essence, needs assessment is the "front-end" of evaluation, and, as Michael Scriven (1978) has said, "it puts the value in evaluation." Evaluation completes the needs assessment by determining the extent to which needs have been met by a program intended to meet those needs.

We know much more about evaluation than we do about needs assessment. It is much easier to assess the effectiveness of a single program than to probe the ill-defined realm of human needs. Evaluation focuses on a single object, while needs assessment focuses on assessing a virtually infinite domain of needs. One of the factors that has led to the relatively slow development of the technology of needs assessment has been the fact that "needs" (real needs) are so difficult to measure. Like attitudes, motives, and personality traits, they are generally hidden below the surface of everyday awareness. It is rarely possible to receive a reliable answer to the question: "What are your needs?"

The most frequently used methods of needs assessment use questionnaires which ask that very question. The problem is that few people are aware of their real needs, or else real needs are confounded by perceptual biases. Often symptoms of needs are confused with real needs. For instance, the person who says he needs more money might only want more money, or money might be a substitute for a lack of self-esteem (or some other internal need). Attempting to discriminate between wants, interests, needs, and other related characteristics has been the most significant thorn in the side of needs assessors for a long time. When educators ask: "What courses do you need?" or "What skills do you need to learn?", it is virtually impossible to distinguish these felt needs from interests and wants. As a result, and to the extent that time and resources permit, conscientious needs assessors are using unobtrusive measures of need, such as archival records, observation of people at work, performance tests, rather than using questionnaires exclusively. The belief is that inferred needs are more legitimate than self-reported needs. However, as long as the limitations of questionnaire, self-report methods are understood, there is no harm in using them as part of a needs assessment effort.

The rest of this paper will deal with some of the other critical issues in needs assessment which are integral to developing an valid and reliable needs assessment instrumentation. Bear in mind that there is no one correct method of needs assessment, and the correct method for each assessor must derive from actual information requirements.

What type of needs are you interested in assessing?

Are there dificiencies in performance, skills, attitudes, motivation? It is vital that the needs assessor have some hypotheses concerning the apparent nature of the need. These hypotheses will determine the most appropriate methods of needs assessment. If there is a performance deficiency, it is important that performance be evaluated. If there is an "interest-need", then a questionnaire might be in order. Hypothesizing about the presumed nature of the need makes it possible to determine a strategy for assessing the need. Without such hypotheses, needs assessment would be like shooting in the dark.

What is the context of the need to be assessed?

Although we all try to be idealistic in needs assessment, and as objective as possible, we must recognize the realities of the situation. An educational organization is looking for needs that can be addressed by educational means, and the needs assessor would not stay employed long if he was not able to identify educational needs for the organization. We should also be aware of the political pressures that may impact on the needs assessment process. Who is powerful in the organization? Who should be involved in the needs assessment process? From the earliest stages of needs assessment, the problems of implementation must be anticipated and addressed. The context of needs assessment should be carefully defined if the needs assessment results are to have any impact at all.

What resources are available for needs assessment and subsequent action?

A fruitful needs assessment must be based on a realistic judgment of resources. We must know early in the process what type of (and how intensive) an effort the organization is willing and able to underwrite. In addition, to what extent is the organization willing and able to satisfy needs that might be assessed? It is the most fundamental rule of needs assessment that there must be commitment to meet with appropriate action any needs that are uncovered. Otherwise, needs assessment will become an exercise in frustration. As Scriven (1978) has explained, needs assessment and action planning to meet anticipated needs should be part and parcel of the same process.

How does one go about developing a plan for needs assessment?

In developing any needs assessment plan it is essential that the following questions be addressed:

1. For whom is the needs assessment intended? Information should be collected in an appropriate form for this person or persons.
2. Who is part of the target population? Although this might seem "Mickey Mouse", a frequent failing in needs assessment is an inadequate awareness of the client population.
3. What data collection methods should be used? The answer to this question depends on the hypothesized nature of the need and resources available.
4. Should sampling techniques be used? It is rarely necessary to observe all members of the target population. Often a small random sample of persons will suffice and decrease cost considerably.
5. Who should be involved and how should they be involved? It is rarely possible to do a comprehensive needs assessment alone, and the help of others in invariably required.

6. What are the anticipated costs and who needs to approve the plan? Required approvals of the plan and costs should be accomplished as soon as possible to facilitate revisions, if necessary, and to secure commitment for the project.

7. What constraints might hinder the needs assessment? Just as it is important to anticipate resources, it is also important to anticipate constraints. This way we can be prepared for the inevitable contingencies which might hinder the project.

How should the information be disseminated and used?

The needs assessor's job does not end with data collection. It is part of his duties to present the information to appropriate decision-makers and make sure that the information is being used appropriately. Those who are not used to making decisions based on needs assessment data might have great difficulty in doing so without help and guidance. The closeness of the needs assessor to the data makes it imperative that he participate in the development of action plans to meet assessed needs. In addition, it is essential that the needs assessor supervise a debriefing to determine the effectiveness of the process as used and to suggest improvements for the future.

Needs assessment is a complex process, the full complexity of which is just beginning to be recognized. It is not just a matter of objective data collection and analysis. This paper has endeavored to explain some of the complexities and suggest methods for dealing with them. It will be some time before the technology of needs assessment is as sophisticated as evaluation technologies. However, it is essential that this be the case, since evaluation is so integrally tied to needs assessment. You see, it is ridiculous to put much credence in the evaluation of a program that might have illegitimate justification for its existence. Many programs, if adequately assessed in terms of needs satisfaction, might never have seen the light of day.

Reference

Scriven, Michael, "Needs Assessment: Concept and Practice," *New Directions for Program Evaluation*, 1, Spring, 1978, pp. 1-11.

ABSTRACT

A visual literacy test was constructed following the _Standards for Educational and Psychological Tests_. Visual literacy was defined and the critical attributes to be tested listed in a Table of Specifications. Visual literacy was shown within the cultural environment which reflects man's orientation to a world of video, and placed in the verbal-visual model represented by left brain-right brain hemispheric research. A pilot test was run. Analysis of this data was used to construct the final form of the test which was given to a stratified random sample of Educational Supervisors, Librarians and Media Technologists in the State of Missouri. Results were reported and a Test Manual and Manual for Administration constructed.

58

Visual Literacy Assessment

A paper presented at the
AECT National Convention
New Orleans, Louisiana

by
Mary Louise Turner
St. Cloud State Univ.

March 8, 1979

Visual Literacy Assessment

I. INTRODUCTION

The mode of communication used in teaching influences learning and retention. In the 1950's P. J. Phillips of The University of Texas suggested some differences in retention related to modalities of learning. Reading was the least efficient while saying and doing or hands-on experience was the most influential. Kinder (1973) quotes Phillips telling us that "Retention of what is learned is likewise related to sense experience. Observation and research tend to show, holding time as nearly constant as possible, that people generally remember:

10 percent of what they read

20 percent of what they hear

30 percent of what they see

50 percent of what they hear and see

70 percent of what they say

90 percent of what they say as they do a thing"

This may be interpreted on a continuum from passive to active involvement in learning or from the symbolic to the actual or real world experience. It further suggests that the closer the facsimile to the real world, the greater the retention of learning.

Ted C. Cobun suggested in 1968 that "we learn about 1 per cent through taste and another 1½ per cent through the sense of touch. The sense of smell provides about 3½ per cent, and hearing provides about 11 per cent of what we learn. An overwhelming 83 per cent of our learning is through visual experiences."

Marshall McLuhan suggested in 1967 that "any understanding of social and cultural change is impossible without a knowledge of the way media work as environments." Television has the capacity to bring us all into one culture, one world with common views and concepts. It could become the great leveler negating real differences between us.

The Trend Report put out quarterly in Washington D.C. forecasts social change. In January of this year it reported major social changes as our country becomes a society whose main product is information. Already fifty percent of our work force is engaged in processing information. By contrast that figure was only 10 percent in 1950. It further suggested that if you are looking for a new career you should try information dissemination.

But this is not a new prediction. In a book written in 1973 and titled Information in 1985 a key forecast suggested that "by 1985 there will be 4 to 7 times as much scientific and technical information as in 1970." Further the demand will only be satisfied by an increasingly accelerated automation which will increase 100-fold over the present systems. Finally, it was suggested that the progress of automation itself may be inadequate to keep up with new information.

So the modes of information dissemination are changing: reading and researching books is giving way on the one hand to computer storage and retrieval of information and on the other hand to video systems. Not only do we have commercial television broadcasts and public television systems but the ability to record off the air programs to be retrieved at will. As yet these are in several formats as in the early stages of audio cassette recording but we are even now thinking in terms of 3D imaging which will add yet another dimension of reality.

This proliferation of information and modalities of dissemination are forcing us to take another look at learning, communication, psychology and sociology. Theories of mass communication assume greater importance than ever before. The hue and cry of parents and PTA's concerned about the influence of sex and violence on television and the effect of advertising on young minds with regard to junk food, etc. is witness to this fact.

It is important to ask ourselves why this is so. As we do so and search the literature for an answer, we find a growing interest in the effects of visuals. Moreover, a concept of visual literacy has been developed during this decade, and research from the physiological arena suggests a tie to differing sensitivites of brain hemispheres: the left side of the brain is sensitive to temporal concepts of order, logic, time, writing, verbalization and reason; the right side of the brain is sensitive to images, space, visualization, invention and intuition.

During the decades that followed the invention of the printing press the left hemisphere attained dominance as learning was almost entirely print oriented and the definition of an educated person related primarily to the ability to read and write using alphabets most of which utilized symbols that in no way reflect what they represent. Visuals on the other hand are pictures of the real world albeit in two dimensions until the discovery of the holographic process using the movement of light through time and space to create 3D images.

We are living in an exciting, changing world which demands changing concepts and understanding of the processes which affect us. One of these processes is the means of acquiring information. School is by no means the only place children acquire information or learning although it is the primary source of formal learning. Television is by far the greatest source of informal learning and preschool children are the single largest television audience in America according to Marie Winn (1977).

"They spend a greater number of total hours and a greater proportion of their waking day watching television than any other group," she says. In 1970, the average was 30.4 hours each week and the figure was going up. By the time a child graduates from high school he or she has spent many more hours in front of a television tube than in school. Why? What is the appeal? Is it important? Should teachers become proficient at utilizing this medium of communication? Should there be or will there be a move toward more and more dissemination of information visually as we move from the age of the printing press to the record and cassette press? What skills will be needed, if indeed this is the case?

We have moved into the realm of non-verbal communication. Many books have been written about body language, the effects of dress upon success, photoanalysis, subliminal influences, embeds, synesthesia and techniques devised or utilized to brainwash us. The media spends huge amounts of money to determine the precise effect of television communication upon specific groups. Motion pictures are now produced following research of acceptability and appeal to the public. Influencing people through the visual mode is becoming a precise science based upon practical research. With a greater involvement of the senses, greater influence can be exerted with longer lasting effects upon retention as suggested earlier.

Jeffrey Schrank (1975) suggests that "we have put vast control over our children's minds into the hands of broadcasters, toy and food manufactures, and other commercial interests whose dominant concern is what's good for profits and sales." Nor is this true of children alone.

Wilson Bryan Key (1972, 1976) suggests that we as adults are being sexually manipulated through subliminal seduction. His books attempt to tell us the secret ways advertising men arouse our desires to sell their products. Nor is this effect only from ads per se. Material is now being incorporated into the entertainment or program section of television to gain the same effects.

This is not a new art. Games of perception and hidden images known as anamorphic art have existed from the renaissance to the present (Leeman, 1975). But today the technology exists to utilize the effects as never before.

Fortunately, we can combat these effects by an awareness of the processes used to influence us. We may also utilize these processes to enhance our communication skills with others. These are some of the reasons for becoming visually literate. But what is involved? What are some of the basics of visual literacy?

Dondis (1973) has taken an artistic approach and recognizes three levels of any visual message: representational, abstract and symbolic. The abstract level refers to the reduction of the visual to its basic visual elements; the symbolic suggests the same sense that print is symbolic: that is, we have attached meaning to specific symbols.

Others have taken this same tack and attempted to define visual communication by its elements. And these are important as any artist knows. However, a perusal of the literature suggests other dimensions such as subliminals, embeds, perspective, illusion, layout, cues, and special attributes of some formats such as film and video tape which allow movement and time manipulation. Some elements have special psychological effects. Color, line and camera angles might be included here. Finally, the ability to operate the equipment and to utilize it maximally is important. As one becomes more adept at equipment synchronization and utilization various combinations of equipment offer added dimensions of communication visually. A good example of effective multi-media utilization occurred during the Bicentennial Year when the Corps of Engineers carried their display to every part of the nation. Most if not all of you saw it. Finally, awareness of techniques such as polarization allow one to extend the capabilities of softwares: slides or transparencies, for example, can show movement. A classic example is a transparency showing the beating of the heart and flow of blood through the body. The heart beat can even be regulated on the spot to match that of any individual watching.

Since all of these concepts affect visual communication in a major way, a Table of Specifications was constructed to include them for the purpose of test construction.

II. PURPOSE AND RATIONALE

The purpose of this study (Turner, 1978) was to construct a visual literacy test and standardize it. Visual literacy was defined and the critical attributes to be tested were chosen after carefully reviewing the literature. These categories were listed in the Table of Specifications which was used as a guide for the test construction. Visual literacy was shown within the cultural environment which reflects man's orientation to a world of video, and placed in the verbal-visual model represented by left brain-right brain hemispheric research. Literature was quoted to support the concept that visual information is more complicated and broader in its definition than verbal information. It is associative. We must become familiar with its elements before moving into a compositional stage. We also need to recognize the psycho-physical forces operating in human perception since visual meaning inheres in all senses and not just in seeing. Whole chunks of information can be assimilated through sight and perception with great speed. Moreover, the control of the final effect of a visual lies in the manipulation of its elements and the planned use of visual techniques.

The concepts of visual literacy chosen for this test were grouped into a 5 x 4 matrix. Within the 5 categories 3 sub-categories were embedded. Visual literacy for the purpose of this test was defined as the ability of people of all ages to read visuals as they read printed materials. It includes the ability to write visually as well using graphics, cameras and other devices. It recognizes that visual literacy, unlike print, is multi-dimensional and perceived through the senses. Words are sometimes cumbersome but a visual can communicate at a glance, arousing in us parts of our own nature, feelings and sensibilities.

The test itself was designed to provide a comprehensive assessment of the ability to communicate visually. It is appropriate for students in high school and college, for teachers, librarians, media specialists, and supervisors. Emphasis is placed upon one's ability to read and construct meaningful visuals.

It is constructed to yield a dependable measure of Spearman's "g." The single score obtained summarizes performance over a wide variety of test materials. There is, however, an emphasis upon visual skills rather than verbal.

We make no distinction between what is being measured and Spearman's "g." The difference is the means of measuring it: that is, the type of item. Intelligence or information of any kind is the same by whatever means of communication unless the means itself is considered to be intelligence. Like Otis-Lennon then we accept the fact that we are measuring one general thing which may be broken into many factors.

Ours is a factor little recognized until the age of electronic media even as print was not a factor until the development of a printing press that could mass produce. Media concentrates information, synthesizes and communicates it in a total sense environment.

Ideation, notation and communication come together in visualization: we need therefore a test to assess more than verbally oriented intelligence measures. The domain used is based upon dimensions mentioned in visual literacy and media research.

The test may be used as a basic competency test for the basic languaging skills of visuals.

Another valid use would be to diagnose individual visual communications problems of teachers, school librarians and media specialists. Unfortunately, even media specialists today have looked too long upon media as hardware instead of a communication medium with all the power, nuances, and cadence of any communication medium. Therefore, they have failed to capitalized upon the available potential of visuals. This test should demonstrate lack of languaging skills within the fifteen dimensions encompassed by the test.

These may be grouped into five factors:

1. non-verbal
2. attending devices
3. theoretical applications

4. meaning of elements

5. media components

Another use of the visual literacy test might be to certify that faculty members possess the basic skills of visual communication.

Still another, to show the public the responsible and accountable manner of teaching the whole student through both verbal and visual communication channels: we are attempting to remedy the limitation of measuring or teaching basic skills via one medium.

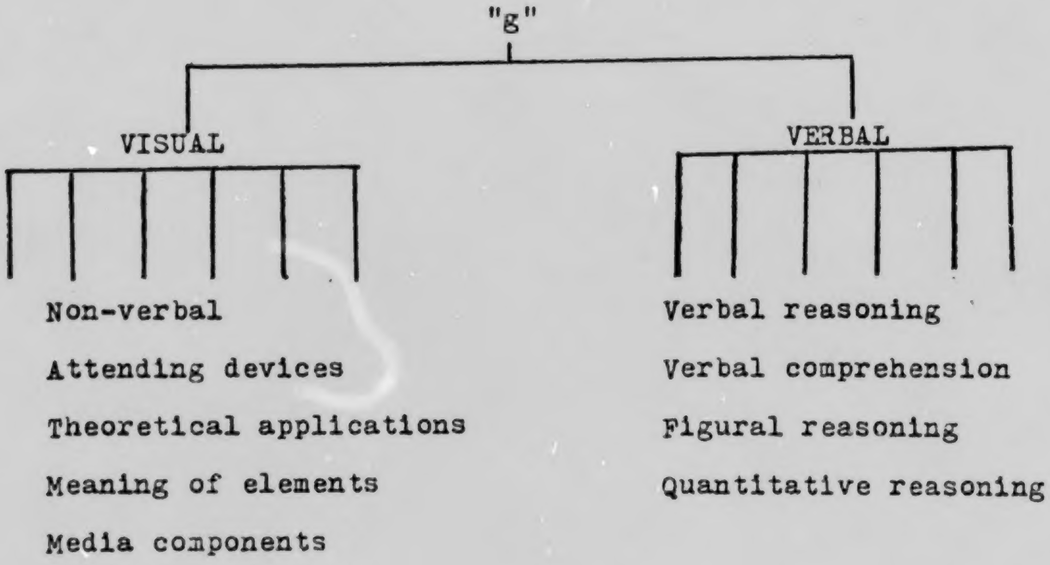

"g"
VISUAL
VERBAL
Non-verbal
Attending devices
Theoretical applications
Meaning of elements
Media components
Verbal reasoning
Verbal comprehension
Figural reasoning
Quantitative reasoning

Figure 1. The division of Spearman's "g" into visual and verbal components.

III. DEVELOPING TEST SPECIFICATIONS

Researchers agree on the importance of the interaction of three main classes of taxonomic variables: media attributes, individual differences, and outcome behaviors. But media attributes are often defined in terms of equipment or operations. Ivor K. Davies (1973) shows that major trends are discernible in the research literature that relates cognitive, affective and psychomotor classes of learning objectives to these categories. However, Richard E. Clark (1975) suggests four possible schemes to describe and select relevant media attributes that interact with indivisual differences: one reasons from extant trait systems like Guilford's structure of the intellect; another from cognitive and/or affective process description; a third from typical or atypical learning outcomes a la Bloom's taxonomy; and the last suggests media researchers should extract media attributes from treatments employed in previous media research. The latest research shows the interrelation of languaging and literacies related to the split-brain syndrome with the use of phonemes and morphemes on the one hand and kinemes and visemes on the other. Larger elements of visual vocabulary include color, shape, line, light, texture, pattern, perspective, etc. There are also ways to express oneself visually, types of visual expression, illusions and visual games which involve learning to see and perceive and are influenced by time, culture, experience and personality. Images that go beyond surface levels involve the sensory registers without which we would have no motion in films or TV. Therefore, the content of this test examines the cognitive/affective aspects of non-verbal communication and influence.

	Identification	Theory	Application	Technology
Non-verbal:				
Body language	2	2		
Subliminals	2	1	1	
Illusions	2		2	
Devices:				
Layout		2	1	1
Format Cues		2	1	1
Attention		2	1	1
Film/Tapes:				
Teach		3		1
Movement		2		2
Time manipulation		1		3
Elements;				
Color		2	2	
Line		2	2	
Camera angle	1	2		1
Media:				
Types	2	2		
Combinations		2	2	
Components	4			

Table 1. Table of Specifications.

Item types: The test uses non-verbal, attending, theoretical, meaning and media component type questions. These include body language, subliminals, and illusions. A second category contains layout, format cues and attention gaining techniques. A third suggests awareness of underlying theoretical applications: the way film teaches, how to create movement and manipulate time. Researcher investigated effects of color, line and camera angle are included as are aspects of media concepts which relate to individual differences.

Levels of difficulty: Item difficulty was held to an average of .5 or 50% to maximize reliability. Any items with discrimination indices greater than .8 or less than .2 were deleted. A spiral format within each quadrant was used: that is, there is a cycle through the types of items four times in the test.

Number of items: There are four questions in each sub-category of five major classifications for a total of sixty questions. This holds the test length to an hour and is adequate to assess visual literacy concepts since each will be presented four times. It is a 5 x 3 x 4 design.

Organization: The test is segmented into four equal quadrants. This means that half of the test could be given in thirty minutes, or a mini-test could be given in 15 or

20 minutes by using one section of the larger test. This allows several ways to analyze the data.

Test items: Visual literacy experts may be asked at a later date to write test items. You are in fact invited to do so. It is planned to develop a large test pool of items to be used in future test construction and development.

Method: The development of the test included the creation of the test specifications, writing the test items, compiling the preliminary test form, trying out the preliminary test form on a sample, pretesting and analyzing the item statistics, compiling the final form, administering the final test form for standardization and preparing the test manual to be used with the test. All of the steps in the construction of the test meet the Standards for Educational and Psychological Tests. A class of media production students was given the preliminary form of the test for item analysis. Using these statistics, the final form of the test was prepared. A random sample from the population of Educational Supervisors, Librarians, Learning Resource and Media Technologists in the elementary and secondary schools of the State of Missouri was used to standardize the test. Data analysis includes mean score, median, standard deviation, semi-interquartile range, Kuder Richardson 20, standard error of measurement and internal consistency. A t-test was performed to determine whether or not there was a significant difference between the two groups. The dependent variable was the test scores.

IV. RESULTS AND DISCUSSION

The pilot test indicated an average item difficulty of .38. By dropping out items with difficulty less than .2, the difficulty level was raised to .53 which met the predetermined standard.

Average discrimination was .2 which is acceptable for classroom tests.

Reliability of the test was .654 which placed it in the typical range of classroom tests. Within categories reliability scores were much higher confirming that we had included several concepts. Non-verbal and devices categories had the lowest correlation internally but still correlations internally exceeded .60. The other categories of film/tape elements and media correlated in the nineties. The results confirmed our expectation that only bits and pieces are known about visual literacy. Knowledge reflected by correctly answered questions related specifically to required course materials. It is our belief that the results of this same test without being changed would differ significantly if given to visual literacy experts. It is hypothesized that there would be no questions with "no correct" answers and, in fact, these and questions showing the lowest correlation on this particular analysis more perfectly query the concept of visual literacy per se.

Form B, a 30 item test, was compiled from information acquired from the pilot test and was given to a stratified random sample of the population of Educational Supervisors, Librarians, Learning Resource and Media Technologists in the State of Missouri with a questionnaire to be used for generalization of test results. The same types of analysis were performed on the data and comparisons between groups tested were made.

Sixty-two percent of the sample responded. This is an unusually high percentage for mail questionnaires. Of the 91 who replied, 16 submitted incomplete data. There were 75 usable tests: 24 were from Educational Supervisors and 51 from Library/Media Professionals. Each district of the State of Missouri was represented by respondents from counties within the districts.

The male/female distribution showed a greater percent of males in the Supervisor group and a greater preponderance of females in the Library/Media group. A t-test showed significant male/female difference at the .1 level but not at the .05 level.

There was a good dispersion of scores. Reliability for the 30 item test was .47 but using the Spearman-Brown prophecy formula to determine reliability for a test of 45 similar items would raise reliability to .59; 60 similar items would raise it to .64; 75 items to .69; and a test of 90 similar test items would increase reliability to .73.

Still another way to increase reliability would be to eliminate or revise the questions with correlations less than .3. There were three of these in our sample. Of course, a better approach may be for subjects to become familiar with these concepts. This, too, would change the item correlation. Comments suggest unfamiliarity with subject matter may be the greatest consideration.

The standard error of measurement is an estimate of the probable extent of error in the test scores and is interpreted in the same manner as a standard deviation. A standard error of measurement of 2.5, for example, suggests that for any particular test score the odds are 2 to 1 that the subject's true score (the average score on several similar tests) would not deviate from the obtained score by more than 2.5 points.

The more reliable and error free a test is, the smaller the error of measurement: that is, the standard deviation of the error distribution. The standard error or measurement for this test is 2.1, a relatively small error of measurement. This suggests that the researcher can place confidence in the accuracy of the test scores. It means that 2 out of 3 times a subject's score in visual literacy will fall within the range of the obtained score on this test ± 2.1 points.

Factors influencing reliability are:

1. the length of the test--the longer the test of similar items, the higher the reliability estimate.

2. less variability in the population: that is, a smaller range of individual differences or talent decreases reliability.

3. unfamiliarity with content also reduces reliability.

4. variability of content--the more similar the items are in content, the greater the intercorrelations, and the higher the reliability estimate. Items with dissimilar content have lower or zero correlations.

5. item difficulty--selection of items of moderate difficulty maximizes total test dispersion. The higher the item dispersion, the higher an item correlates with other items. However, it is more important to select similar items than it is to control item difficulty to maximize reliability.

Item analysis of discrimination and difficulty were again run. For a detailed explanation see Turner, 1978, pages 111 through 114.

V. VALIDITY

No test review is complete without information about its validity: that is, the degree to which a test measures what it is suppose to measure. Cattell (1964) suggested that test validation appears in a continuum from practical to conceptual validity: that is, from criterion to a focus on the test itself. Purpose defines the method of validation to be used. "With content validity," says Lemke (1976), "the concepts to be measured are specified by the test developer." This was done in the Table of Specifications provided by the researcher. Lemke continues that content validity is provided by a logical analysis of the domain of items as was done with the pilot test, with the final form of the test being constructed according to item analysis statistics from the pilot test.

Construct validity, on the other hand, is concerned with traits embedded in the test and the degree to which these account for performance. Correlation analysis can be used to establish the construct validity of a test. Unfortunately, no other visual literacy test exists to date against which this test can be correlated. To this researcher's knowledge, no standards have been established for mediated instructional expertise against which this test might be compared. The field of visual literacy itself is new having developed within this decade.

The other type of validation with which the researcher might be interested is criterion validity. This type of validity can be established by expectancy tables and simple correlations with appropriate variables when these exist. Since visual literacy is a new field, content validity seemed the more appropriate. This is especially true in light of the stand by the National Educational Association with regard to all tests being a type of achievement test. It is recognized by test construction experts that content validity is most appropriate for achievement tests. Moreover, criterion validity takes time to develop since it is based upon predictive statistics.

Reliability is a necessary, but insufficient, condition for criterion validity. Moreover, the maximum value of the validity coefficient is the square root of the reliability coefficient. The correction for attenuation estimates the correlation between true scores of the test and the criterion and allows an estimation of validity if the measurement instrument is perfectly reliable.

VI. RECOMMENDATIONS

It was shown that a visual literacy test can be constructed to measure visual communication skills. Some areas of visual literacy are more familiar to educators than others. Scores tend to be lower when subjects are unfamiliar with visual literacy or visual communication skills. It seems evident that interest in visual literacy will increase. This is suggested by the number of persons asking to use the test and comments about it. Finally, there is a relationship between visual literacy expertise and media education. It is suggested that the item pool be expanded and norming data enlarged.

Specifically, results suggest:

1. a viable visual literacy test can be constructed.

2. some areas of visual literacy are more familiar to educators than others.

3. scores tend to be low because of unfamiliarity with the subject of visual literacy or communication

4. test reliability could be increased by dividing the test into subtests with the major categories listed in The Table of Specifications as guides. Reliability could then be figured for each subtest instead of the test as a whole.

5. reliability would probably be increased by

giving the test to a non-specialized group with a wider range of individual differences with respect to media.

6. interest in visual literacy will increase--this was evident from comments about interest in media and the request for permission to give the test to a High School Media Studies Class.

7. there is a relationship between visual literacy expertise and media education.

REFERENCES

Anderla, G. Information in 1985. Organisation for Economic Co-operation and Development, 1973.

Cattell, R. Validity and reliability: a proposed more basic set of concepts. Journal of Educational Psychology, 1964, 55, 16-22.

Clark, R. E. Constructing a taxonomy of media attributes for research purposes. Audiovisual Communication Review, 1975, 23(2), 197-215.

Cobun, T. C. Media and public school communications. In R. A. Weisgerber (Ed.). Instructional process and media innovation. Chicago: Rand McNally & Co., 1968.

Davies, I. K. Competency based learning: technology, management and design. New York: McGraw-Hill, 1973.

Dondis, D. A. A primer of visual literacy. Cambridge, Mass.: The M.I.T. Press, 1973.

Key, B. W. Media sexploitation. New American Library, 1976.

Key, B. W. Subliminal seduction. New York: Prentice-Hall, 1972.

Kinder, J. S. Using instructional media. New York: Van Nostrand Co., 1973.

Leeman, F. Hidden images. New York: Abrahms, 1975.

Lemke, E. & Wiersma, W. Principles of psychological measurement. Chicago: Rand McNally College Pub. Co., 1976.

McLuhan, M. Understanding media. New York: McGraw-Hill, 1964.

Schrank, J. Understanding mass media. Skokie, Ill,: National Textbook Co., 1975.

Turner, M. L. Visual literacy assessment of mediated instructional expertise for educators. Unpublished doctoral dissertation, University of Missouri, 1978.

FIELD-TESTING INSTRUCTIONAL MATERIALS

by

Bill Winn
University of Calgary

Paper presented at the annual convention of the Association for Educational Communications and Technology, New Orleans, March 1979.

Abstract

The application of research and theory to the development of instructional materials relies largely for its success upon testing materials in the field and revising them accordingly. This need not necessarily be a complicated business. The practitioner who is not a trained instructional developer, can test the effectiveness, feasibility and the affective component of instruction. This paper describes a few simple methods for doing so.

In 1972, AECT's Committee on Definition and Terminology published a definition of educational technology, (1972). The basis of this definition was a statement by John Kenneth Galbraith, from *The New Industrial State*, that "Technology is the systematic application of scientific or other organized knowledge to practical tasks", (1967, p.12). The three components of this statement, systematic application, organized knowledge, and practical tasks, describe the relationship that exists between research on the one hand, and instructional development on the other, between the theory and the practice of educational technology. AECT's definition can be interpreted to mean that educational technology is a three-phase problem-solving process whereby educational problems are identified, appropriate knowledge is marshalled that is likely to provide a solution to these problems, and whereby this knowledge is systematically applied so that the solution is both useable and effective. Instructional development is pivotal in this process.

If we look a little more closely at each of the three components of Galbraith's statement, the nature of the development process, and the relationships between the various roles of the educational technologist become clearer. We must first ask what are the practical tasks of educational technologists. Within the limited context of the activities of researchers and instructional developers, the answer to this question must be: to bring about situations in which learning can effectively occur. In practice, this means that the instructional developer is told of instructional need that has to be met. A need is usually taken to be "a measurable discrepancy between what is and what should be" (Kaufman, 1972); so the role of the instructional developer becomes one of bridging a gap, with theory, research and experience as his bridging materials. To take a simple example, if a school board discovers that its high-school graduates cannot read, then a need has been identified: they are illiterate (what is); they should be able to read novels, magazines, and newspapers without difficulty (what should be). This gap is measurable insofar as such factors as working vocabulary size, reading speed and reading habits can be assessed empirically. The instructional developer's problem, then, is to develop instructional materials, strategies, environments which make the need disappear.

One place the instructional developer can turn for a solution to problems is to the body of research and theory that has been accumulated over the years. Here may be found information about how developers solved similar problems elsewhere, which can be applied directly, or more general psychological theories of how people read, why students can get through school and still be illiterate, and the best ways of teaching reading. The developer must also look at the type of learner in question. Are they motivated? Are they of high, average or low ability? The feasibility of the various strategies that begin to emerge as possible ways of solving the problem must also be taken into account. The developer, in effect, is gathering a large amount of information about possible solutions to the problem and about the particular context in which the problem has presented itself. This information gathering, through learner analysis, task analysis, feasibility analysis, and the provision of several potential solutions, is a critical step. Solutions to instructional problems can never be better than the information on which they were based, and they are frequently worse.

It is in the systematic application of this information to a particular problem that the instructional developer must excel. The application of theory to practical problems is systematic if it meets two criteria: first, the theory, experience and research that are applied must be pertinent; second, the decisions arising from the application of particular research and theory must be empirically tested. The second point is extremely important. It does not matter how "correct" or relevant research and theory appear to be. They will never be perfectly appropriate to the particular circumstances in which the instructional developer has to apply an instructional solution. After all, tests of statistical significance are never made at the 100 percent level. And the generalizability, that characterizes theory, implies the need for adaptation to specific situations. So however sure the instructional developer is that the right instructional decisions have been made, the instructional products must nevertheless be tried out in the field and revised until predetermined standards are achieved at which point the developer will be sure that the need has been met. The rest of this paper addresses itself to this question.

Evaluation

Before getting into the "how to" of field testing, it is necessary to dwell for a moment on the more general question of evaluation. It is customary to distinguish between two general types of evaluation: formative and summative (Bloom, Hastings and Madaus, 1971). The difference between the two is described by Gagné and Briggs (1974) in terms of what each type is used for. In their words, formative evaluation is used to "form" instruction; summative evaluation is used to determine the "summed" effects of instruction on learners. The main purpose of formative evaluation is to "debug" early versions of instructional materials and activities, to see if they work, to identify weaknesses, to find out why they do not work, with a view to using the information gathered by evaluation to revise the materials and activities in ways that correct those problems. Information produced by formative evaluation is prescriptive. The purpose of summative evaluation is to gather information that is descriptive. Summative evaluation is carried out on larger units of instruction that are in some sense complete. The information gathered tells us how the finished package performs in normal instructional settings, and is often used to validate instruction (Thiagarajan, 1971). This paper is concerned with formative evaluation.

So that it is useful to the instructional developer (this term applies to anyone who is developing instructional materials, not necessarily a professional developer), formative evaluation of instruction must provide information on the effectiveness, feasibility and affective qualities of the instruction under development. While there is a tendency for field testing to get complicated, each of these three kinds of information can be gathered easily. Yet even simple methods must allow the developer to do more than simply conclude that the instruction does not work, cannot be implemented, or that the teachers and students do not like it. The assessment of the effectiveness of instruction under development must not only evaluate how much students learn, but must also permit the developer to pinpoint what the student does not learn well, and why. The purpose of formative evaluation, remember, is to guide revision of materials. It is therefore important to know not just that revisions are needed, but where they are needed. Likewise, formative evaluation must identify any reasons for infeasibility of implementation. If teachers say that they cannot manage to carry out their part in the proposed instructional process, the developer needs to know why so that feasible alternatives can be tried next time around. Similarly, the affective component of formative evaluation must determine why students like some parts and not others, which parts are boring and which are exciting, which are interesting and which dull. Only then can the portions learners like less be made more palatable.

Evaluation in each of these three areas is only practical if a level of acceptable performance for the instruction is determined beforehand. After all, even the best designed instruction can never be one hundred percent perfect. It would therefore be a waste of the developer's time to keep on revising materials until all students scored full marks, or all teachers could take part, or everyone liked them. It is safe to say that the earliest revisions are the easiest to make since they deal with the most obvious errors and consequently will bring about the most marked improvements. Each subsequent round of field testing and revision will require increasingly finer adjustments to be made until a point is reached where the improvement in instruction brought about by the revision is so small that it is not to be worth the effort (and cost) to make it. While the developer can usually tell when this point has been reached, it is of considerable benefit to try to anticipate when the point of balance between what is effective and what is worthwhile will be arrived at. This is done by setting up criteria for acceptability. For example, it might be decided that no further revisions will be carried out when all learners score eighty percent or more on a posttest, or when a simple majority of teachers judge the new instruction to be feasible, or when no more than ten percent of the learners say that they do not like what has been developed. Sometimes these criteria are set by the developer's client, as in the case of performance contracting. But this need not concern us now. Suffice it to say that criteria, set beforehand, often help the developer make the decision that further revisions are not necessary.

The rest of this paper describes a few simple yet perfectly adequate techniques for field-testing instructional materials. They are not intended for fully-fledged instructional developers, who have their own repertoire of far more sophisticated tests and methods. They are intended, rather, for the eduational practitioner, who is not a trained or experienced developer, who has neither the time nor the money to carry out elaborate evaluation studies, and whose development activities are modest in scope. The methods to be described have been gathered from far and wide, and for the most part used successfully by staff in the Learning Technology Unit at the University of Calgary (Learning Technology Unit, 1979).

Field Testing Instructional Effectiveness

Criteria. Typical criteria for the acceptance of instruction are:

i. Percentage performance on a posttest, ranging from a low of 70% to a high of 90% depending on the task or content.

ii. Number of students making errors, which is particularly useful for debugging instruction in motor skills.

iii. Time taken for students to complete a task on which 100% success is expected, such, again, as performing simple motor skills, or straightforward recall.

iv. Percentage of students "passing" a posttest, where the pass mark could be 100%, or could vary as in i.

v. Estimates of creativity, originality, success at problem-solving for higher-level learning (Gagné, 1970; Bloom, 1956).

Task Description. The developer must identify what discreet tasks the learner must be able to perform in order to achieve overall success. This breaking down of instruction into its sub-parts must be carried out for all types of instruction. It is wrong to think of the word "task" as applying just to motor skills. Cognitive content can be broken down into intellectual tasks. The task description provides both a framework around which the developer builds instructional materials and activities, and also a check-list that is used in field-testing.

For psychomotor skills, a task description can be carried out by thinking through the task, and writing down on a sheet of paper every step that has to be completed. It is best to err on the side of being too detailed at this point. It is easier to remove steps from the list than to add them later. Better still is to have a person who can do the task go through it, while writing down everything that the person does. This way it is less easy to overlook steps when you think it through. For cognitive skills, working through the task yourself and consulting others are about the best you can do to compile your list of sub-tasks. You cannot see what is going on inside someone elses head (though this can be deduced fairly accurately in some instances, such as solving mathematical problems, or translating into a foreign language, since each step or word produces something on paper). More complicated and empirical methods exist for the analysis of cognitive tasks; but they are best left alone by all but the most highly trained developers.

The result of the task description is a list of smaller tasks in sequence. These sub-tasks form the basis for the evaluation of the effectiveness of instruction.

Testing. To test the effectiveness of instruction in motor skills, the developer must actually observe the learner working through the series of tasks, and note on a score sheet whether each step is successfully completed or not. Learners should work through the instructional material and associated activities, and should then be tested individually on the skill. All that it is necessary to record is whether the learner was successful or not in completing the task. (A typical checklist for testing a motor skill -- laminating a picture -- is given in figure one).

Figure 1 About Here

If the learner becomes completely stuck, the examiner can help so that the learner can move onto the next step in the sequence. A note should be made of the examiner's intervention.

Testing cognitive skills will usually involve a written test to be completed after instruction. Remember, this is formative, not summative evaluation. It is therefore necessary to test not just the cumulative effects of instruction on the learner's knowledge and ability, but the successful acquisition of each sub-part of the total skill as it is encountered along the way. As a general rule, there should therefore be at least one question on each cognitive sub-task identified in the task description. For example, if the objective of the instructional unit is for the learner to be able to describe why an airplane flies, and the task description has identified the understanding of vector diagrams and of airflow over a curved surface as tasks leading to an understanding of the whole process, then the learner should be asked at least one question on vector diagrams and one on airflow as well as general questions on how planes fly. The reason for this is that if you just had a learner describe how planes fly, and the learner was unable to do so, then you would not know if it was because there was a lack of understanding of vector diagrams or of airflow. You would know that instruction had not been successful, but you would not know where the learner had had difficulty.

Tests of cognitive abilities may be of any kind -- multiple choice, short answer, or even essay, provided that instructions are given to the learner which require an explanation of all pertinent sub-tasks. A word of warning, though. It is difficult to construct good multiple-choice tests, which are valid, reliable and discriminate well between those who have learned the material and those who have not.

Sample Size. You are field-testing, not conducting a sophisticated psychological study. The early rounds of field-testing should identify the most serious weaknesses and the most obvious mistakes. These should become apparent in the behavior and in the written answers of the first few of your learners. On some psycho-motor skills, it is not necessary to test more than three or four learners in the first round of field-testing, increasing the number to half a dozen or as many as eight in subsequent rounds. The same is true for cognitive skills. Written answers from four learners will give you a satisfactory first estimate of the success of your instructional materials. Keeping the numbers small will also leave sufficient learners in the pool of potential "subjects" for subsequent rounds of testing and evaluation.

Interpreting Scores. There is no need to submit learners' scores to elaborate statistical analysis. Looking at individual learners' scores is quite sufficient to identify where problems lie. Figure two shows a score-sheet of errors committed by just two learners during field testing

Figure 2 About Here

of the laminating unit mentioned above. Each learner made two attempts at the task. Both missed step eight the first time around (cutting off excess laminating film), and one of them became quite confused over the whole sequence for preparing the film to cover the illustration during the second trial. This should be enough to indicate to the developer that the instruction was not clear about removing excess film and preparing the film to lay over the illustration. Testing another fifty learners would not add much more information about this.

The interpretation of scores on written answers of whatever kind works in the same way. Incorrect answers to questions based on sub-skills upon which an understanding of the whole is based point to where the instruction is not as effective as it might be. The inability of more than one learner out of four to answer correctly a question on vector diagrams would suggest to the developer of the unit on flight that instruction was weak at this point.

Revision. The developer now looks at the faulty sections of the instructional materials that have been identified in the testing. Revisions can be referred back to any point in the development process. Perhaps the section of videotape showing excess laminating film being cut off was terribly out of focus, and just needs to be re-shot. Perhaps the producer did not think it was terribly important, and showed it far too quickly to be understood. Perhaps it is not really necessary to cut off excess film at this point. Or, perhaps the understanding of vector diagrams is asking too much of the learners, and the objective is not attainable. Whatever revisions are needed, the developer should make them, and submit the materials and activities to a second round of testing. The test-revise cycle is continued until the pre-established criteria have been met.

Evaluating the Feasibility of Instruction

Criteria. Typical criteria for assessing the feasibility of implementing instructional materials and activities are:

i. Cost. Can the "client" afford to purchase and maintain the instructional system?

ii. Available equipment and facilities. Normally, cost and available facilities are known before development begins, so these constraints would be taken into account before field-testing.

iii. Personnel. Are the teachers who will be using the instructional materials and activities trained to use them?

iv. Cost-effectiveness. Maybe learning takes place and the client can afford the system. But it is still relatively expensive for the amount of learning that takes place.

v. Difficulties in carrying out instruction. Maybe the instruction is successful, but requires an inordinate amount of time, or preparation.

Evaluation. The evaluation of the instructional materials and activities against these criteria can be done informally. Cost and facility factors can be assessed quite simply by questioning teachers and administrators. Personnel skills and difficulties in carrying out the instruction can be estimated by direct observation of teachers and learners trying the materials out in classrooms. Maybe teachers need inservice training to implement the instruction, and maybe the activities and materials should be organized in different ways so that they are easier to carry out and use. The observer in the classroom should take notes of any logistical or other problems that occur so that modifications can be made to overcome them.

The above stresses the importance of doing at least some of your field testing in classroom settings if it is at all possible. The testing of individual learners, described in the previous section, needs to be supplemented by try-outs under conditions as close as possible to those in which instruction will be implemented. It is only through this kind of observation that certain problems can be detected. A set of self-instructional units, produced by the Learning Technology Unit at the University of Calgary (Learning Technology Unit, 1979) ran into a problem which observers noted. Materials designed for use by one learner were in fact being worked through by several learners at a time. As a result of this a division of labor occurred and no one student performed all of the required tasks. The reason for this was a higher enrolment in a lab. component of a course than had been anticipated, and the problem was overcome by duplicating stations in the lab. area. The infeasibility of self-instruction under these circumstances might have gone undetected had not part of the field-testing involved direct observation in the classroom setting.

Evaluating the Affective Component

Criteria. Instruction should reach acceptable levels in the following areas:

i. The students should like the activities and materials. If twenty percent or more indicate dislike, revisions should be made.

ii. The materials should capture and hold the interest of the learners. If twenty percent or more indicate boredom, the materials should be made more interesting.

iii. The materials should provide encouragement (feedback, reinforcement) to the students. If they feel a lack of encouragement or of direction, revisions should be made.

Testing Affect. More is going to be said in this session about measuring attitudes. So let us look at a few simple ways of finding out learners' affective reactions to instruction. Urie Bronfenbrenner (1976) has pointed out that the simplest way to find out if learners like the instruction, (or experimental treatments) they receive is to ask them. Every developer should at least do this. As learners run through the instructional materials and are tested, it is not difficult to ask them a few questions, such as: "Did you find this interesting or boring?" "Was it too easy or too difficult?" "Would you like to be taught more lessons this way?" Responses to these informal questions can be noted for future reference. Interest can be increased in revisions by speeding up, or even just changing the pace of instruction; by switching from a verbal to a visual form of presentation, or vice versa; by inserting an example into an expository segment of instruction, and so on.

Further refinements to these simple techniques can be made. Information can be obtained by having each learner, after working through the materials, rate statements on Likert scales. Typical statements would be: "I found the material interesting," rated from "strongly agree" to "strongly disagree". Information about the learner's likes and dislikes can, in this way, be quantified, and sometimes this is an advantage. Or an affective component can be built into an observer's checklist (Thiagarajan, 1971). Here, in addition to a check mark or a cross against each step, to indicate success or failure, other symbols can be used. An exclamation point might indicate boredom, an asterisk, enthusiasm. In this way, revisions can be directed to specific portions of the instructional materials and activities, and motivational devices can be built into the parts of the materials where they are most needed.

It is also important to test the affective reactions of the teachers, or of whoever it is who will be present when the materials are used. The students may learn what they are supposed to, and they may like the materials very much. But if the teachers do not approve of the way the learners are to be taught, or if they are not enthusiastic about what the developer has produced, then it is likely that success will be difficult to achieve. Again, teachers can simply be asked their opinion, or can rate statements on scales. The developer needs no more information than this.

Summary

A Bare Minimum. For someone who is not accustomed to doing instructional development, and who is faced with the task of developing and producing materials and activities for a particular purpose, the foregoing has provided some suggestions as to how the developer's product can be tried out. The purpose of that trial is to provide information that will permit sensible revisions to be made to materials and activities with a minimum of fuss. To achieve this, it is not necessary for the casual developer to carry out an elaborate, empirical, statistical evaulation. But the developer should at least carry out the following:

i. Set criteria for acceptable learner performance.

ii. Determine criteria for feasibility of implementation.

iii. Determine what will be an acceptable affective rating.

iv. Perform a task description.

v. Prepare an observer's checklist, or written test, depending on the content, (motor skill or cognitive), to evaluate the effectiveness of the instruction for each step as well as for overall ability.

iv. After giving the test, or observing performance, ask the learners if they liked the instruction, if they found it hard or easy.

vii. Quiz teachers and administrators on the feasibility of implementing the instruction.

viii. Ask teachers if they liked the instruction, and if they think it is appropriate for the intended learners.

ix. Carry out revisions, after a try-out with four to six learners.

x. Try to field-test the instruction, or some part of it, in a real classroom setting.

Useful Elaborations. For those with a little more time and initiative, a few additional features can be added to field-testing which make the task easier, more empirical and more objective:

xi. If you have skill in the matter, write multiple-choice tests so that the distractors perform a diagnostic function, i.e. b. is the right answer, but chosing c. indicates that the learner has a different kind of difficulty from the learners who chose a.

xii. Develop a coding system on observation checklists to indicate reactions, such as boredom and enthusiasm, beyond just right and wrong.

xiii. Use rating scales for affective responses.

With these few simple tools, any teacher or anyone who is not a trained instructional developer, should be able to carry out simple field-testing of instruction and revise it accordingly. While development of large and complex instructional systems requires far more elaborate field-testing, and while even field-testing of simpler systems and materials can be more complicated, educational technology is not inherently esoteric. Galbraith's definition of technology as the application of knowledge to practical tasks is not difficult to understand, nor to apply. And it is in the application, through field-testing, of theory to practice, that educational practitioners can, in their everyday activities, become educational technologists.

References

Bloom, B.S. *Taxonomy of Educational Objectives. Handbook I: Cognitive Domain*. New York: McKay, 1956.

Bloom, B.S., Hastings, J.T. and Madaus, G.F. *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill, 1971.

Bronfenbrenner, U. The experimental ecology of education. *Educational Researcher*, 1976, 5, 5-15.

Committee on Definition and Terminology. The field of educational technology: A statement of definition. *Audiovisual Instruction*, 1972, 17 (October), 36-43.

Gagné, R.M. *The Conditions of Learning*. New York: Holt, Rinehart and Winston, 1970.

Gagné, R.M., and Briggs, L.T. *Principles of Instructional Design*. New York: Holt, Rinehart and Winston, 1974.

Galbraith, J.K. *The New Industrial State*. Boston: Houghton-Mifflin, 1967.

Kaufman, R.A. *Educational System Planning*. Engelwood Cliffs: Prentice-Hall, 1972.

Learning Technology Unit. *Evaluation of Practicum Self-Instructional Units*. (L.T.U. Research Report No. 2). Calgary: Learning Technology Unit, 1979.

Thiagarajan, S. *The Programming Process: A Practical Guide*. Worthington Ohio: C.A. Jones, 1971.

FIGURE I

DRY MOUNT PRESS: Task Description
/SEALAMIN

TASK	YES	NO
1. Set thermostat at 225° F.		
2. Turn on		
3. Set tacking iron to 'medium'		
4. Select appropriate illustration		
5. Select Sealamin large enough to cover		
6. Dull side down		
7. Smooth out Sealamin		
8. Cut off excess Sealamin		
9. Place materials in newsprint sandwich		
10. Place sandwich in press		
11. Materials in for 30 seconds		
12. Check for bubbles, wrinkles		
13. (Reinsert if wrinkled)		
14. (Final trim)		

FIGURE II

MATERIALS PRODUCTION LAB

Tasks	Trial #	Task Errors with Two Subjects 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Dry Mounting	1			1		1	1			1				2						
	2									1				1						
Chartex	1			1						1										
	2							1		1										
Spirit Duplication	1														1					
	2					1	1													
Laminating (Sealamin)	1								2											
	2						1	1	1	1		1								
Laminating (Contact paper)	1							2												
	2		1							1	1	1	1	1						

NOTE: For task descriptions see appendices

PART V:

PROJECT REPORTS AND EVALUATIONS

OUTCOMES OF THE INSTRUCTIONAL SYSTEMS DEVELOPMENT PROJECT

by
Robert K. Branson
Center for Educational Technology
Learning Systems Institute
Florida State University
Tallahassee, Florida

Prepared for Presentation at
American Educational Research Association Convention
March 5-9, 1979
New Orleans, Louisiana

OUTCOMES OF THE INSTRUCTIONAL SYSTEMS DEVELOPMENT PROJECT

Robert K. Branson
Florida State University

ABSTRACT

This project sought to meet documented Army training needs through application of system analytic procedures. Doctrine, regulations, and school organization were investigated. Regulations were clear and comprehensive. School organization lacked a clear training evaluation component. During the project, the Army reorganized the schools to reflect requirements of the regulations and to permit proper execution of the new training doctrine. This change greatly increased the role of evaluation in the schools. Recommendations were made about organization and regulations and recommended doctrine was published in detail as the Interservice Procedures for Instructional Systems Development. Lessons learned and management recommendations are presented.

OUTCOMES OF THE INSTRUCTIONAL SYSTEMS DEVELOPMENT PROJECT

Robert K. Branson
Florida State University

Orientation

The session for which this paper was prepared had as the central theme "Lessons Learned" in major instructional development projects. This paper will begin with the broader focus of human resources development, then concentrate on instruction as one of the major approaches to human resources development. By limiting the scope principally to training programs, it will be possible to highlight specific issues not always apparent in typical education settings.

The distinction between education and training is made here simply on the basis of the probability that the position or job classification following training is known prior to an individual's beginning the training program. For example, consider the baccalaureates in chemistry. The probability of a good prediction of their future employment when they enter college is low. At the other extreme, a sailor who enters the Navy's radioman school has a high probability of going to a radioman's job. Instruction, then, for the radioman would be training, for the chemistry major, education. Not because of the subject matter but because of the predictability of the job for which the individual was being prepared. Obviously, in this sense, instruction for which there was no occupational outcome regardless of how specific would be cast under the rubric of education. Dentistry is a training program, business is questionable, while teachers are trained, history majors, education, etc.

The Project

The project to be described in this paper was undertaken by the Center for Educational Technology (CET) at Florida State University. The Center was organized by Robert M. Morgan, Robert M. Gagné, Leslie J. Briggs, and the author in the early 1970's for the purpose of taking a systems analytic approach to the solution of human resources development problems nationally and internationally. Here, a systems analytic approach refers to the application of psychological and human factors research and development techniques to the solution of man-environment performance problems. Each of the individuals noted has detailed his own view of how such applications can best be made by presenting either theoretical treatises or the description of large-scale projects. To gain a better perspective of how these authors view the analytic approach, see, for example, Gagné 1966, 1977; Briggs, 1977; Morgan and Chadwick, 1971; Morgan 1979; and Branson, et al., 1975.

The U.S. Army Combat Arms Training Board (CATB), a highly effective and creative military training organization, approached the Center in 1973 to assist them in making a major revision and upgrading of the approach to Army training. One request made of CET was to assess the state of the art in empirically designed training technology, to identify concepts appropriate to the mission of Army schools and training centers and to recommend methods for institutionalizing these concepts within the Army school and training center system. Careful human engineering of the organization of guidance material and doctrine was essential, particularly in providing "how to do it" guidance to military and civilian training specialists within the Army school system (Branson, Stone, Hannum, and Rayner, 1973).

A second request was to employ appropriate development procedures to prepare and validate a training model with detailed guidance procedures to develop in Army personnel the skills necessary to enable them to analyze, design, develop, implement and control empirically designed instruction. It was my privilege, and I might add, a considerable professional challenge, to have been selected as the principal investigator on that project.

The CET Approach

Principally because of the orientation of the founders of the Center and based on the experience gained from a series of progressively more effective projects, a somewhat traditional systems oriented human resources development approach was taken. A brief description of that approach would be useful. It is well founded in the psychological literature that there are normally only three ways to achieve or improve the performance of individuals in their jobs. One can select people who already know how to do a job. This approach is commonly taken by commercial airlines when they hire former military pilots who have already logged thousands of hours in multi-engine service.

A second approach is taken through the provision of sufficient instruction to individuals so that they will be able to perform. In a military context, literally thousands of people are trained every day to repair vehicles and aircraft, to troubleshoot electronic systems, and so on; people who knew nothing about those systems prior to their entry into the military service.

A third approach to improving or achieving performance is through effective management and supervision. Such approaches would include the manipulation of economic incentives for sales people, for example, or through work enrichment procedures to create more favorable working conditions and assignments for people on the job. Generally, business and industrial organizations pay far more attention to the management aspects of performance than they do to the selection and training aspects of performance.

Mager and Pipe (1970) have detailed these three approaches in a straightforward fashion. It is their view that before any effective human resources change can be made, it is first necessary to document a clear performance discrepancy. That is, to define an outcome gap between what is and what ought to be. Some of those outcome gaps will be caused by faulty selection procedures, others will be caused by skill deficiencies, while others will be caused by inadequate approaches to management. Training can only be a solution for performance discrepancies that are documented to be skill deficiences.

It is important to recognize that all human resources development programs or projects will be effective only when they are viewed as a means to the achievement of some known outcome. While that should be obvious, frequently it has been found that a large number of people view training as an end in itself rather than a means to achieve an acceptable level of performance. There is abundant literature detailing situations in which training was used as a solution to a non-training problem, and was therefore used ineffectively. One reason for the apparent failure of effective training programs is that they represent the most innocuous intervention management can take.

The processes of documenting specific outcome gaps have been detailed by Kaufman (1976). The processes of needs analysis and assessment are undertaken to establish program validity. In a training program, validity refers to the correlation between job requirements and the instruction. An important first step in defining outcome gaps is the identification of an organization's mission. An organization should first be evaluated only against its own mission. That is important because there is often disagreement among outsiders on what an organization's mission ought to be or should be, rather than what it actually is. Kaufman also argued that it is our obligation to question an organization's mission, but, once established, the mission then becomes the guidance for the entire organization.

Once the organization's mission has been established it has been rather a straightforward matter to compare the operations of the organization to its mission including the way that the enterprise is organized, the resources available, and the use of those resources in the accomplishment of the mission. It is a fundamental requirement that results or outcomes be an important part of the total assessment procedure. Prior history has shown that we far too often get concerned with process rather than outcome. These outcomes or results must then be compared to plans or expectations. Finally the quantitative or qualitative discrepancy between results and plans becomes the operational definition of the outcome gap.

The Politics of Discrepancies

To a seasoned systems analyst, the discovery and documentation of an outcome gap represents an opportunity to apply a variety of solution alternatives. Unfortunately, to the manager of an agency or a corporate executive, the documentation of an outcome gap often becomes a political problem of considerable proportion. On the one hand it is impossible to go forward with precision in the absence of documented outcome gaps. On the other hand the existence of a documented outcome gap often requires action on the part of the executives responsible. These actions are not always solution oriented.

When the documentation of an outcome gap is done for the purpose of establishing a political advantage, it is often the executive's tactic to discredit those who prepared the documentation. For example, in February, 1979, the General Accounting Office issued a highly critical report of Peace Corps recruiting activities, citing as evidence that the Peace Corps had "failed to interview" adequately a large number of those hired. This "discrepancy" was cited as an example of poor management on the part of the Peace Corps. Yet, fifty years of psychological research on interviewing has demonstrated that the interview is the least effective commonly employed method of employee selection.

A second approach taken to the handling of discrepancies is often that of finding someone to be the scapegoat. Whose fault is it that the training program is not working? Let's find the people responsible and replace or discipline them. This decisive executive action creates the illusion that the problem has somehow been solved when in fact it will recur. Having effectively scapegoated the problem, however, the executive can remain in place until the ultimate tragedy occurs. Scapegoating often protects individual jobs but rarely contributes to mission.

Generating Solution Alternatives

Sometimes it is possible to document discrepancies in an organization for which no one can reasonably be scapegoated. The responsibility for the problem is so diffuse that no one can be clearly blamed. It was the fortunate case in the military ISD project to have been able to document the discrepancy in such a way that making it public did not require administrative or punitive action. Once opportunities of that kind are discovered it is possible to begin constructive deliberations on solution alternatives.

Locating Contributors

It was decided early in the military Instructional Systems Development (ISD) project to encourage wide participation as a means of generating a base of participation and support. First, we visited several Army installations to identify knowledgeable and effective staff memebers. Each of the Army's schools and training centers has a number of civilians and military personnel who are competent in the applications of a systems approach to training (SAT). The participation of these specialists in the development of the ISD materials was critical. First, they have experienced a wide range of prior training research and development activities. Second, they could serve as a responsive editorial board for the critique of any tentatively selected approach. Many of them could make particularly good estimates of what will and what will not work in various environments.

Selected people from those installations visited were invited to attend various planning and briefing sessions during the planning and design of the project. Accumulating a group of experts or specialists to address a problem is not without risks. Many participants had already made up their minds about "the" correct approach before they arrived. Fortunately, others adopted a problem solving orientation and were willing to consider a variety of solution alternatives. Those who had strongly entrenched views-whose views did not prevail--either became detractors from the project or dropped out all together. Perhaps some day we'll be able to work out an effective method to get such groups to remain in a problem solving mode rather than a solution selling mode.

It should be noted here that the Army has worked out over the years a variety of methods for handling decision making. One of those and probably the most prevalent is that of having the senior officer present make the final decision on a project or program. Until the time that the decision is made, people are free to argue, to present alternatives, to oppose others and so forth. However, after the decision is made it is expected that all others participating will execute the action plan with diligence. But, even in the Army, the better decisions are more faithfully executed than those which are suboptimal.

The Army State-of-the-Art in 1973

In addition to identifying the key people who were at the various training centers and schools in the Army, a second purpose of our visits to those installations was to define the existing state of the art. Because the Army had done the early basic research in the systems approach to training, it was decided that a clear picture of exactly where they were in 1973 would be basic to making any real improvements. Accordingly, we analyzed the training, training literature, school organization, and training of trainers status as of the first quarter of 1973.

Three important areas were analyzed and then evaluated in terms of the mission they were intended to accomplish. The first of these areas was that of finding out what the people were supposed to do and how they were supposed to do it. In the Army what people are supposed to do is spelled out in regulations, and how they are supposed to do it is spelled out in other forms of guidance. The regulation which specified what and how the Army was to train in 1973 was called the Systems Engineering of

Training, Regulation 350-100-1. That regulation, issued in 1968, had as its basis a fairly standard systems approach to training model. That model is displayed in Figure 1.

A second consideration was how people were prepared to do what they were supposed to do. Because there were a large number of people in the Army school system concerned with training and an even larger number of people attending Army schools, there must be considerable and regular attention paid to how people are prepared to their jobs. Third, after people have been prepared to do their jobs then it is necessary to have some regular and objective way of finding out how well they did their job.

The results of the preliminary studies and visits left us with a good understanding of what the Army schools were supposed to do, how the people at the schools were prepared to do their work, and how they were evaluated and informed of the quality of their work. The last element of that analysis was that of understanding how they were organized to accomplish the assigned work.

In the Systems Engineering of Training model displayed in Figure 1, the first step or function called for is that of job analysis. It follows according to management logic that if job analysis is a function required to be performed then someone should be assigned to perform that function. Further, we should be able to find on the organizational chart where in the organization that that function is performed. In this sense the organizational chart is a road map which allows us to match function to organization.

Displayed in Figure 2 is the organizational chart that was presented to us by schools we visited during 1973. We were able to track the function of job analysis down through the Deputy Commandant for Doctrine through

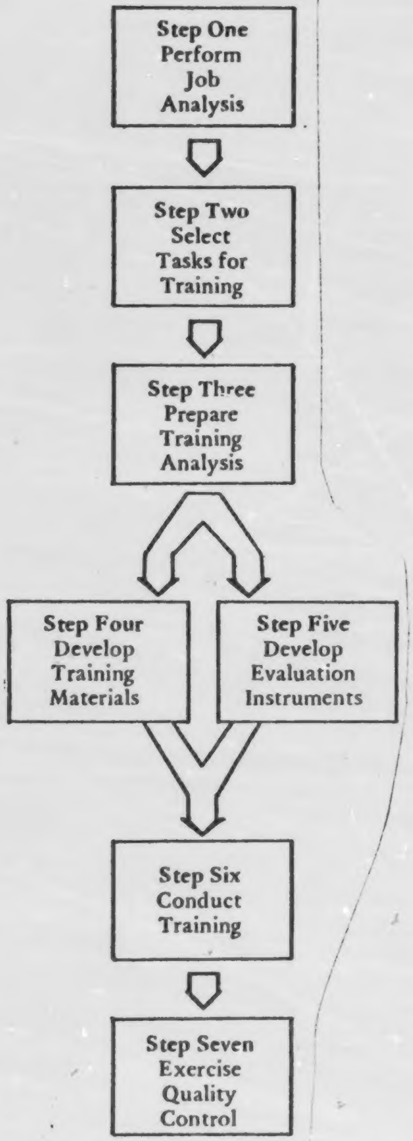

Step One
Perform
Job
Analysis
Step Two
Select
Tasks for
Training
Step Three
Prepare
Training
Analysis
Step Four
Develop
Training
Materials
Step Five
Develop
Evaluation
Instruments
Step Six
Conduct
Training
Step Seven
Exercise
Quality
Control

Figure 1. TRADOC Regulation 350-100-1 Training Model

SCHOOLS MODEL

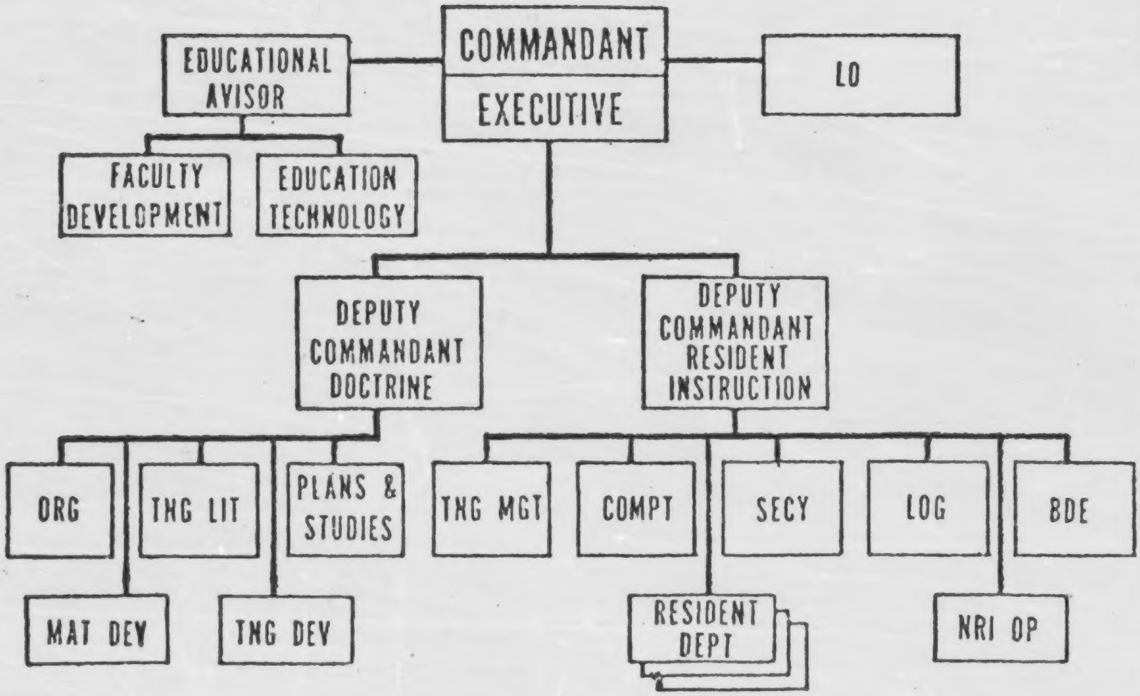

COMMANDANT
EXECUTIVE
EDUCATIONAL AVISOR
LO
FACULTY DEVELOPMENT
EDUCATION TECHNOLOGY
DEPUTY COMMANDANT DOCTRINE
DEPUTY COMMANDANT RESIDENT INSTRUCTION
ORG
TNG LIT
PLANS & STUDIES
TNG MGT
COMPT
SECY
LOG
BDE
MAT DEV
TNG DEV
RESIDENT DEPT
NRI OP

Figure 2. U.S. Army School's 1973 Organizational Chart

the Directorate of Training Development and found that organization charged with the job analysis function.

Subsequently, we tracked each of the other functions specified in the regulation down through the organizational chart until we could find the person who was actually in charge of conducting the work specified in the regulation. Each time we found a discrepancy between what was called for in the regulation and what was actually being done, either organizationally or functionally within the school, we documented that discrepancy.

Putting the -ing in Training

Fundamental to the theory of a systems approach is the notion of control. Control refers to the process of planning, executing the plan, obtaining results, comparing the results to the plan, and acting on the discrepancies. Because there are many elements of a training system, identifying the method of control becomes a key aspect of analyzing the effectiveness of the organization.

Consider as an example an accounting department. In the accounting department, accountants apply standard procedures and techniques to each of the accounts on a periodic basis and present periodic reports. The purpose and form of those reports is specified in advance and the person in charge of accounting makes certain that each of the accountants applies the specified procedures in a craftsmanlike manner. Individual accountants are not free to handle their accounts according to their own whim. Further, any deviations from standard practice are documented and explained. Not only are accounts managed, and accountants managed, but the entire process of accounting is managed to meet specifications and, accordingly, is controlled.

Applying that model to training would require that specified plans be made for the trainees, that the trainers would conduct training according to specification and, that the results of training would be compared to those plans, discrepancies documented, and action initiated on those discrepancies.

In the Army school's model, we were unable to identify any evidence that training as such was being controlled. Trainers were being trained and trainees were being trained by trainers, but the results of that training were not compared to the plan by anyone specifically charged with that responsibility. While it may be true that people in various departments had interest in certain operations data about training, no one was charged with the responsibility for documenting training results.

Based on this analysis of the regulation and the organization of the schools, we documented this major organizational deficiency in a consultants' report to the president of CATB in 1973.

Our interviews with training specialists in the field identified many who claimed that the regulation under which they were operating was not a good regulation. A consultant group, HumRRO, was hired to investigate the application of the regulation at the Army aviation school. While that report contained many useful recommendations which were followed in the ISD project regarding ways to improve the regulation itself and its documentation and evaluation procedures, no mention was made of the fact that the school was not organized to implement the regulation (Ricketson, Schulz & Wright, 1970).

One viewpoint from which the ISD project analysis was made included the following basic tenant: If a function, result, task, or outcome is everyone's responsibility, it will not get done. Applying this rule to the

Army's 1973 school organization, the evaluation and control of training was widely dispersed in the organization and as a consequence did not occur in a meaningful and deliberate way.

While there are no "laws" which apply universally in training management, and there is no such thing as the correct organization, a number of reliable principles have emerged over the years which work more often than not. We have generally approached new situations with these principles in mind. The first six laws of training management:

1. Systems, models, designs, and equipment do not achieve results. Managers do.
2. If a function, result, task, or outcome is everyone's responsibility, it will not get done.

   Corollary: A manager can only handle N-1 collateral duties. (One less than the number already assigned.)
3. Any activity which is not thoroughly checked, or inspected according to specification or requirement will not be done according to specification or requirement.
4. Incumbents in any job (mechanics or instructors) do not want to learn how to do it our way--they want to do it their way, no matter how historically ineffective.
5. An enterprise must be adequately organized and resourced to accomplish mission.

   Corollary: Training and education are resource sinks; no quantity of resources will ever be perceived as adequate.
6. To be effective, a management systems analyst must pick a problem to work on which is big enough to be important but for which no one can reasonably be blamed.

Our conclusion in 1973 was that the regulation was perfectly adequate for the intended purpose. There was insufficient guidance, examples, and training provided for people to follow the regulation but those are peripheral issues to the question of the quality of the regulation. Thus, if the regulation has all of the attributes of an SAT model, there is little point in revising. What was required was detailed guidance, training, and manager training.

At that point in the project, we tried to summarize the data that had been gathered so that we could test the quality of our inferences about the status quo. It was important that these inferences be as accurate as possible since they would serve as the basis for all future plans.

In that context, it occurred to us that placing the blame on the regulation was a clever rationalization on the part of the managers. Perhaps not even a conscious one. What we found was, to us, a clear violation of the first law of training management: Systems, procedures, models, and machines do not produre reulsts; managers do.

The violation was not that the managers were not working hard or that they were ignoring the duties, it was that no specific training results were required and no one was responsible for achieving (or evaluating) those results. Results in this context refers solely to outcomes, not processes. The Army was certainly diligent in applying the training process.

An effective solution would require not only a means for specifying and measuring outcomes, but also, an organizational recognition that a senior manager must be held accountable for the achievement of those outcomes. To achieve an organizational recognition of those outcomes would require a major reorganization of the schools. That reorganization was clearly beyond the scope of work for which we had been contracted, and more importantly, well beyond the decision authority of our sponsoring agency. Consequently, we were not optimistic about a fundamental change.

The Development Plan

By early summer of 1973 we had completed the analysis of the state-of-the-art in training in the Army (Task I) and, with the assistance of a number of training experts, had been able to outline the project plan. This plan included the development of an Army training model (to replace the one in the systems engineering regulation), development of a detailed set of procedures to provide training guidance to users, technical level workshops, workshops for the immediate supervisors of technical work, and workshops for senior managers. The workshops for the supervisors and managers were thought to be the most critical feature of the project. This management emphasis was based on the assumption that supervisors cannot readily achieve results they do not understand.

In addition to the specifications for completing the project, a methodology was established in which each of the major program elements would be developed, critiqued, and thoroughly staffed before proceeding to the next step. While this process of printing and distributing draft materials was time consuming, it was absolutely essential. The process of assimiliating and interpreting the critique and suggestions was difficult at best. Often, we found ourselves torn between two equally effective alternative approaches, being able to use only one of them for consistency purposes. Finally, we projected a requirement to use the materials for a period of 18 months, visit installations where they were being used, conduct an in-depth evaluation of the use of the materials, and then do a final version based on the formative data accumulated during the tryout of the draft. Unfortunately, that revision did not occur.

It is difficult to be sure why the revision did not take place according to schedule. Some of the reasons centered around the change in project officer and a change in organization of CATB. There was also a small group of powerful civilians called "education advisers," whose jobs we had recommended to be eliminated or restructured during the Task I study, and there was a change in TRADOC at the highest levels. Perhaps the contractor was simply ineffective in convincing the Army that the final version was necessary--they still issue the draft version.

Mission Drift

Hardly a month had passed following the award of the contract for the development of an Army model and set of Army procedures when we encountered an organization called the Interservice Training Review Organization (ITRO). A subcommittee of that organization was at the time attempting to standardize training terminology among the services. For those of you in the university community, this task was akin to that of trying to get several schools within a state university system to agree on common courses. It always sounds so logical on the face of it until one actually tries to execute a plan.

Upon hearing of our project, that ITRO committee approached the Army with the proposal that the model and procedures being developed for them be converted to an interservice sponsorship, that is, the project would be supervised by respresentative members of the Army, Navy, Air Force and Marine Corps. In the interest of interservice harmony, the Army agreed to that proposal.

While it seems reasonable that it ought to be possible to prepare training materials which would be useful to all members of the armed forces, that reasonableness applies only to principles, rules, generalizations, concepts and the like. It does not work at all when one attempts to describe training problems, training procedures, sources and uses of data, and other service specific requirements. Clearly, our mission had begun to drift. It is often difficult to please everyone. Fortunately, at that time, the project manager assigned by the Army remained the same. As all defense contractors are aware, high quality projects cannot be completed if high quality project managers are not assigned to complete them.

It is the duty of the project manager to keep the contractor clearly oriented toward the mission and objectives. Such a responsibility is particularly important when it is necessary to work with a large number of people in order to complete the project. Often, working with a large number of people permits the introduction of noise into the system. The government project manager must taken into account the legitimate suggestions and concerns of those participating while, at the same time, keeping the project on track so that the outcome will be in their best interests.

Late in the project, we went through the rather shocking experience of a change of project manager. The first project manager was a highly rated lieutenant colonel in the regular Army. When he was reassigned overseas, he was replaced by an energetic, technically competent captain in the Utah National Guard. Imagine your all-American high school quarterback suddenly being sent to the Dallas Cowboys. Captains, no matter how brillant, do not vote as much stock as lieutenant colonels.

We believe the lesson learned here is that it is not in the best interest of the contractor to change project managers in the middle of the project unless one can make a substantial change for the better. We believe this lesson would generalize across a wide range of projects.

Results

The outcomes of the project can be described in two major categories: Those which were expected and those which were not. The major expected outcome of the project was the publication of all of the draft materials developed. These materials included:

- Interservice ISD Model (See Figure 3)
- The manuals published in the Army as TRADOC Pam 350-30 and in the Navy as NAVEDTRA 106A
- The workshop materials, including the exercises and the workshop director's guide
- A slide/tape presentation which provides an overview of the entire set of procedures.

All of the materials described above are referenced at the end of this paper with their document acquisition numbers so that interested civilians can find them accessable.

A significant element in the Task I problem analysis was that of the organizational structure of the Army school system. Perhaps the most dramatic unexpected outcome during the entire project was the reorganization of the Army school system. Following the publication of the manuals (in late 1975), the Army issued a draft organizational chart for the school systems in 1976. This revised organization is referred to as School Model '76, is presented as Figure 4. Of particular interest, in Figure 4 is the Directorate of Evaluation which has been placed on

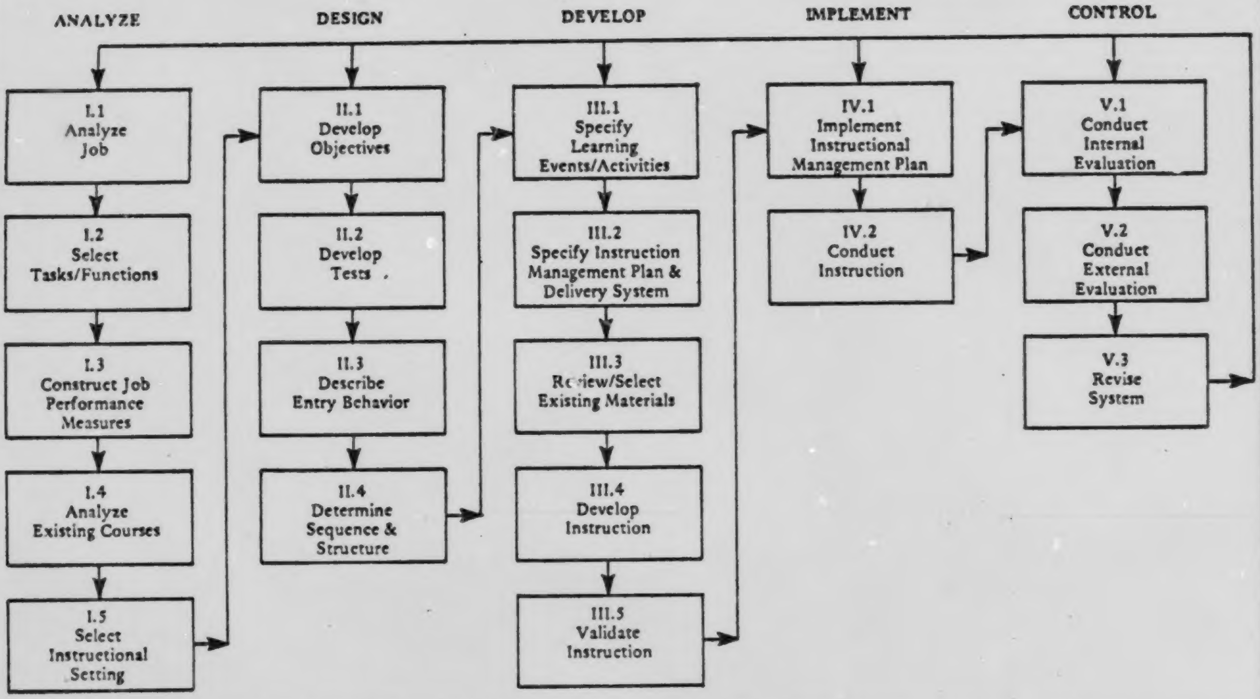
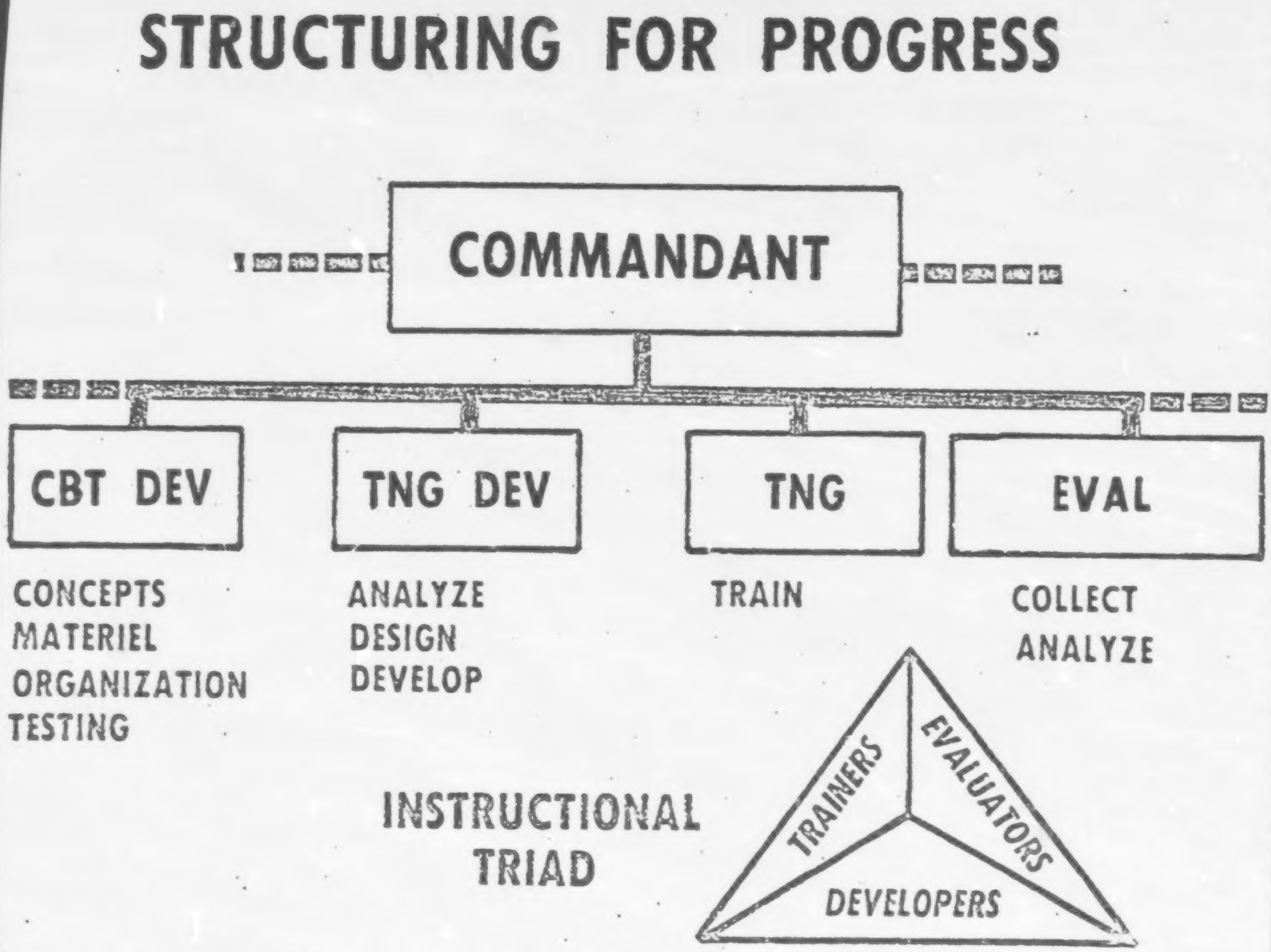

ANALYZE
DESIGN
DEVELOP
IMPLEMENT
CONTROL
I.1 Analyze Job
I.2 Select Tasks/Functions
I.3 Construct Job Performance Measures
I.4 Analyze Existing Courses
I.5 Select Instructional Setting
II.1 Develop Objectives
II.2 Develop Tests
II.3 Describe Entry Behavior
II.4 Determine Sequence & Structure
III.1 Specify Learning Events/Activities
III.2 Specify Instruction Management Plan & Delivery System
III.3 Review/Select Existing Materials
III.4 Develop Instruction
III.5 Validate Instruction
IV.1 Implement Instructional Management Plan
IV.2 Conduct Instruction
V.1 Conduct Internal Evaluation
V.2 Conduct External Evaluation
V.3 Revise System

Figure 3. Interservice Procedures for Instructional Systems Development Model (TRADOC Pam 350-30)



STRUCTURING FOR PROGRESS

COMMANDANT

CBT DEV	TNG DEV	TNG	EVAL
CONCEPTS MATERIEL ORGANIZATION TESTING	ANALYZE DESIGN DEVELOP	TRAIN	COLLECT ANALYZE

INSTRUCTIONAL
TRIAD

TRAINERS
EVALUATORS
DEVELOPERS

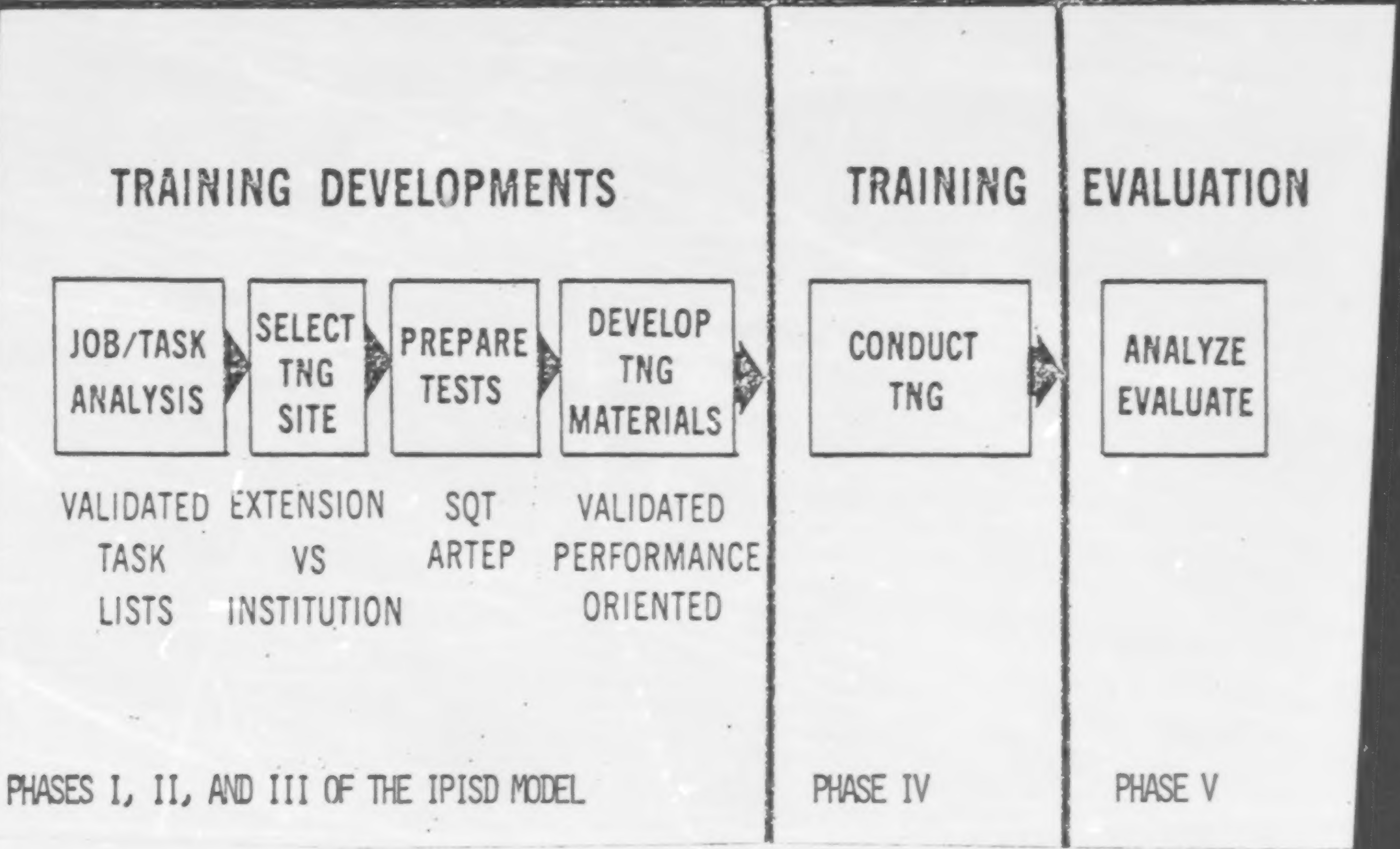
Figure 4. U.S. Army School Model 1976

an equal level with all other major divisions of the organization. What this means is that the Directorate of Evaluation, which previously did not exist, now is on the same level as the Directorate of Training Development, the Directorate of Training, and the Directorate of Combat Development. Now, the people who develop the content of training work in a different department than those who develop the training and training materials. Those who conduct the training are in a different department from those who evaluate the outcomes of training. This independence of major function provides a system of checks and balances which permits feedback to be channeled to appropriate elements of the organization as a means of improving outcomes, and the potential for gaining control.

Figure 5 shows how the relationship between the organizational departments relates to the phases of the ISD model. The Directorate of Training Development is responsible for Phases I, II, and III of the ISD Model. The Training Directorate is responsible for Phase IV of the ISD Model. And Phase V, Evaluation, is placed in the Evaluation Directorate. The School Model '76 represents a considerable conceptual improvement over the prior organization since it placed equal emphasis on the evaluation of outcomes as it does with the conduct of training processes. If this training organization can be kept in place long enough for its systematic attributes to mature and operate, it will have been a major contribution to school management.

A second unexpected outcome was that the final version of the materials was not completed. A thorough revision was planned based on broad and sufficient usage data in the Army. The revision was to have been similar to a revision of instructional materials based on formative evaluation

SYSTEMS APPROACH
TO TRAINING



TRAINING DEVELOPMENTS

JOB/TASK ANALYSIS

SELECT TNG SITE

PREPARE TESTS

DEVELOP TNG MATERIALS

VALIDATED TASK LISTS

EXTENSION VS INSTITUTION

SQT ARTEP

VALIDATED PERFORMANCE ORIENTED

PHASES I, II, AND III OF THE IPISD MODEL

TRAINING

CONDUCT TNG

PHASE IV

EVALUATION

ANALYZE EVALUATE

PHASE V

Figure 5. Relationship between Organization Departments and Phases of the ISD Model

data. As a consequence, many rough edges (and perhaps some huge potholes) remain in the materials. Such revisions cannot be competently done by people not thoroughly familiar with the principles upon which the procedures are based.

What we fear, and there is evidence to support this fear, is that the potholes and rough edges will be identified by some users. These users will try to remove the portions of the doctrine offensive to them, even though that section or step was a fundamental part of the principles being established. If this practice continues, the integrity of the entire model can be degraded.

The third unexpected outcome was the degree of resistance among several civilian education advisers within the school system. One the one hand, they were unable to muster really significant support for their resistance, but on the other hand, they did create a substantial block to implementation. Fortunately, their protest was ill-timed. Had they put forth their maximum effort at the time the project director was changed, or at other times of partial confusion, they could have increased their chances to prevent the publication of the materials.

References

Branson, R. K., Rayner, G. T., Cox, J. L., Furman, J. P., King, F J, & Hannum, W. H. *Interservice procedures for instructional systems development* (5 vols.) (TRADOC Pam 350-30). Ft. Monroe, Va.: U.S. Army Training and Doctrine Command, August 1975. (NTIS Nos. AD-A019 486 - AD-A019 490).

Branson, R. K., Stone, J. H., Hannum, W. H., & Rayner, G. T. *Analysis and assessment of the state-of-the-art in instructional technology* (Final Report: Task I). Ft. Benning, Ga.: U.S. Army Combat Arms Training Board, 1973. (NTIS No. AD-A019 394).

Briggs, L. J. (Ed.) *Instructional design: Principles and applications*. Englewood Cliffs, N. J.: Educational Technology Publications, 1977.

Department of the Army. *Systems engineering of training* (CON Regulation 350-100-1). Ft. Monroe, Va.: Headquarters, U. S. Continental Army Command, February 1968.

Gagné, R. M. (Ed.). *Psychological principles in systems development*. New York: Holt, Rinehart and Winston, 1966.

Gagné, R. M. *The conditions of learning* (3rd ed.). New York: Holt, Rinehart and Winston, 1977.

Kaufman, R. A. *Needs assessment*. San Diego: University Consortium for Instructional Development and Technology, 1976.

Mager, R., & Pipe, P. *Analyzing performance problems*. Belmont, Ca.: Fearon Publishers, 1970.

Morgan, R. M. *The Korean Education Development Institute: Its organization and function*. (Report for UNESCO) New York: United Nations. Educational Scientific and Cultural Organization, January 1979.

Morgan, R. M., & Chadwick, C. B. (Eds.). *Systems analysis for educational change: The Republic of Korea* (Final Report Contract AID/ea-120). Tallahassee, Fl.: Department of Educational Research, Florida State University, April 1971. (Available from University of Florida Press, Gainesville, Fl.)

Ricketson, D. S., Schulz, R. E., & Wright, R. H. *Review of the CONARC systems engineering of training program and its implementation at the United States Army Aviation School*. Alexandria, Va.: Human Resources Research Organization (HumRRO), April 1970.

The Design, Development and Evaluation of an Evaluative Computer Simulation

Presented at the Association for Educational Communication
and Technology National Convention in New Orleans 1979

Research and Theory Division

by

Lisa R. Ehrlich, M. A.
University of Iowa College of Medicine

The Design, Development and Evaluation of an Evaluative Computer Simulation

Lisa R. Ehrlich, M. A.

University of Iowa College of Medicine

This presentation will first compare and contrast the differing purposes of instructional and evaluative computer based simulations. It will go on to discuss the development of a cardiology simulation that will be used as a referent for discussion. The design and structure of this referent simulation will then be presented. The major portion of this paper will discuss the evaluation design of this cardiology simulation and some measurement issues of computer based evaluative simulation.

The technique of computer based simulation has been successfully utilized in an instructional mode for some time in many academic disciplines. Within the past two decades, much preliminary work has been conducted to investigate the use of simulation techniques, on and off line, as evaluative instruments. While the technique of constructing a model of reality for simulation purposes remains essentially the same for either an evaluative or an instructional simulation, the instructional designs differ vastly. This paper will first discuss the differences in instructional design necessitated by the differing educational purposes of these two types of computer based simulations. Once these differences are established, then development and evaluation issues will be presented and discussed using an ongoing research project at the University of Iowa College of Medicine as a model reference. The final portion of this paper will discuss considerations relevant to scoring of evaluative simulations and the issues of validity and reliability.

Purposes of Instructional and Evaluative Simulations

The purpose of an instructional computer based simulation is to allow the student to interact with a simulated model and to learn how this simulated enviornment reacts to differing conditions imposed on it. Most instructional simulations have as part of their design some instructional segments, or, programs that help to guide and teach the learner various principles or conditions present in this simulated enviornment. Many of these instructional segments are tutorial in that they teach the student some of the salient features of the model, or simulation, and allow the learner to progress from very basic principles, or concepts, to more complex aspects of the system of interest. The more sophisticated instructional simulations utilize remedial branching as part of their design to reinforce learning. Some of these complex instructional simulations have internal test items that are used as a check for learner comprehension and understanding of the concepts presented as part of the tutorial package.

Once the learner has successfully grasped the basic principles and concepts, s/he is first taught how to interact with the model, or simulated environment, and ultimately allowed to experiment and manipulate this model. In this phase of the instructional simulation the learner sets up the varying conditions and constraints and then explores how the model reacts given these determinants.

To summarize, the general design of instructional computer based simulations is to first create a model of the object system, to develop a series of instructional tutorial programs which teach various concepts and principles of the object system and to ultimately allow the learner to interact with this model by manipulateing and varying the conditions. The primary purpose of this type of computer based simulation is instructional.

By contrast, the primary purpose of an evaluative simulation is to test, or measure knowledge, skills, and/or competencies within a specific discipline or subject area. Its purpose is not instructional; any learning that does occur is secondary to the main purpose of evaluation. The same principle holds here as with any written cognitive test instrument: whatever a student learns while being tested is purely secondary to the purpose of evaluating what a student "knows" or "understands".

Development of a Cardiology Evaluative Simulation

As stated earlier, the evaluative simulation which will be used as a reference is one developed at the University of Iowa College of Medicine in Cardiology. The purpose of this computer based evaluative simulation is to assess both diagnostic and therapeutic skills of primary care physicians and medical students on acute myocardial infarction. The design of the simulation allows for assessment of diagnostic skills of physicians in sections of medical history taking, physical examination and in the ordering of diagnostic laboratory tests. Patient management skills are evaluated for patients who have uncomplicated illness - in this case myocardial infarction, as well as in section dealing with common complications of the disease: ventricular arrhythmias, bradycardia, hypertension, cardiogenic shock, and cardiac arrest and resusciation.

In June of 1977 a development team consisting of a content expert (cardiologist), a medical liason (physician's assistant and/or a senior medical student), and an instructional designer-programmer began a two year project to develop an on line evaluative simulation in cardiology. A review of the literature was first done to ascertain the state of the art. Once this was accomplished, the content expert and the instructional designer met to begin the content development of this simulation. These "brain storming" sessions lead to an overall design of the "simulated patient", or environment, which was then initially flowcharted. At this point a senior medical student was brought in and introduced to the concept and technique of evaluative simulations and flowcharting. For the first six months, many meeting occured between the three members of the development team involving discussion of the case and the ramifications of diagnosis and treatment. The medical men had no computer or instuctional design background and the instructional designer had no medical background, so, each member of the team learned what exact information was necessary for the development of each part of the simulation.

Since the major portion of this presentation is to center around measurement considerations of evaluative simulations, the design and development aspects will not be detailed. Suffice it to say that the design, development, programming, and the debugging processes of the simulation took one and a half years to accomplish. The programming was done in BASIC on a Hewlett-Packard 2000 Access minicomputer located at the University of Iowa Computer Assisted Instruction Laboratory.

Design and Structure of the Cardiology Simulation

Prior to consideration of some measurement issues, a short discussion of doctor (user) interaction with the simulation model is warranted. Since the purpose of this instrument is evaluative, the user is essentially allowed to progress through the simulation at her/his own rate and with minimal structure imposed. The program was designed to allow the doctor to follow whichever path(s) s/he chooses and is consequently forced to deal with whatever complications arise. Therefore, there is no direction or guidance imposed.

There is no instructional remediation incorporated within this simulation, since instruction is not the intent of this instrument. The feedback the doctor receives is twofold: one form is the result(s) of treatment(s) that were ordered and how this treatment(s) affected the patients' condition. The purpose of this feedback is primarily informational rather than instructional. The second form of feedback is both informational and instructional. This feedback occurs either at the end of the simulation when the patient is successfully treated and released from the hospital, or, if during the simulation the doctor, through incorrect diagnosis or treatment kills the "patient". The program records which sections the doctor has completed and once the simulation is terminated, successfully or unsuccessfully, gives detailed feedback in terms of diagnosis, recommended treatment and rationale. This feedback is included as a summary of the case.

The structure of the simulation is as follows: the simulation is broken down into an introductory office scenario, and a series of management programs on premature ventricular contractons (PVC), bradycardia (day 2), hypertension (day 3), and a final program on cardiogenic shock (day 4). There are also two different cardio pulmonary resuscition (CPR) programs; one for management of cardiac arrest in the coranary care unit and the other for treatment of cardiac arrest in an office or unmonitored ward. There is continuity of presentation of these different management programs. The doctor will proceed through the simulation without realizing that these programs constitute different sections.

The actual logistics of this sequencing will not be addressed in detail as it is complex and not the primary intent of this paper. It is important to note that at many points throughout the simulation the patient can die due to improper management. Consequently, some of the doctors will not get to all the sections of this simulaiton if the patient dies prematurely.

Validity Issues of Computer Based Evaluative Simulation

With the purpose and structure of this simulation in mind, attention is now turned to various measurement considerations and the research design that will be used in evaluating this simulation. It should be noted that the measurement issues that will be raised are not specific to this particular simulation, but, generalize to most evaluative simulations on and off line.

The first step in the evaluation process is validation. Therefore, the primary purpose of the research design is to validate the simulation. The audiences that will be used are not themselves being evaluated. It is their performances on the simulation and on various external validation instruments that are of utmost concern in ascertaining the validity and the reliability of this simulation.

There are four different audiences that will be used to validate this evaluative simulation. These audiences consist of a group of senior medical students at the University of Iowa, a group of primary care physicians who are returning to the University of Iowa for a five day workshop as part of a continuing medical education program, the staff cardiologists at the University, and 20 to 30 internists located throughout the state of Iowa. A general overview of the validating instruments will be discussed as well as the research designs for each group. The data handling and statistical analysis will not be presented since they are technical rather than conceptual issues.

There will be three external validating instruments used: 1) a cognitive written multiple choice exam 2) a subjective clinical rating scale and 3) a medical audit.

The cognitive written multiple choice exam will be used with the senior medical student group and the group of primary care physicians. This exam will contain questions which address the doctors' knowledge level of cardiology and cardiovascular disease. There will also be a series of questions used which address more skill related areas of cardiology. These questions will force the doctor to go beyond the knowledge level, to apply and analyze specific circumstances in order to diagnose and manage patients in various situations.

This exam will be tailored to correspond to the major management points addressed in the simulation. By constructing the exam in this manner, it will enable direct comparison of specific performances on the simulation and on related test items. It is felt that this is the only means of using a cognitive exam as a validating instrument. To help validate the simulation, the scores obtained on the simulation will be compared with the exam results to assess 1) if there is a relationship between scores on each part of the exam and performance on the corresponding segments of the simulation and 2) if the results on the total simulation breakdown into the same relative rankings as they did on the exam.

The subjective clinical rating scale will only be used with the medical student group. The major problems in using this subjective scale as a validating instrument are inter and intra rater reliabilities, as well as the actual observation time each rater has with each student. There is unfortunitly no way to regulate the amount of time a rater has with each student, as most of these encounters are spontaneous. The consistancy with which a rater rates individual students is a problem with any observational instrument and an attempt will be made to control for this. This instrument will be constructed in a manner to enable the students to be rated on specific diagnosis and mangement criterion that corresponds to major diagnostic and management points in the simulation.

The ultimate criterion for validating the simulation is to assess how well actions and decisions taken on the simulation match actions taken in real life. There exists in medicine a means by which this assessment can be accomplished, that is by conducting a medical audit. A patients' chart contains all the relevant information on the diagnosis and management of a case.

It is with this record of information that a case can be reconstructed and the major medical decisions determined. A medical audit is the means by which a case is reconstructed.

Twenty to thirty internists throughout the state of Iowa will be participating in this phase of the validation. With their permission an audit will be conducted on those patients' charts that directly relate to the content of the simulation. Specific comparisons will be made between performance in real life, as assessed through the audit, and the diagnosis and management of the simulated patient.

The criterion related, or more specifically, the concurrent validity of this simulation will be assessed by using the cognitive written test instrument. The relative rankings on the exam will be compared with the relative rankings for both groups on the simulation.

It is assumed that this written cognitive instrument is both valid and reliable and is in fact measuring the same skills that the simulation is measuring. This is a critical assumption. A simulation does not duplicate reality, it only imitates it. Because of this, in assessing the concurrent validity of this simulation it is only possible to determine some aspects of physician performance. This, or any simulation can only give a measure of how a physician is capable of behaving not how well s/he will behave in the actual situation. It is important to bear this in mind when determining the concurrent validity of simulations.

The clinical ratings of the student group will also be used to assess concurrent validity. This too will use relative rankings of the students on individual sections of the clinical rating scale. The results on specific sections of the simulation will be compared with the ratings. It is hoped that there will prove to be a relationship between how the students perform in the clinic and how they perform on the simulation.

Finally, the results of the medical audit will be compared with the physician results on the simulation. It is hoped that this will be the ultimate validating criterion. If it is found that the physicians' performance on the simulation relates to their performance in practice, then some proof of concurrent validity can be established.

It seems a logical construct to assume that a physician who has both diagnosed and managed patients in both clinical and office situations will perform "better" on the simulation than medical students who have had limited experience. It also seems logical to assume that cardiologists should perform "better" on the simulation then either of the above groups.

Construct validity will be determined by comparing the performance of these differing groups. Construct validity will be tested as follows: 1) the primary care physicians should perform "better" as a group on the simulation than the group of medical students and 2) the Cardiologists should perform "better" as a group on the simulation then either the physician or student group. If the simulation is able to differentiate the performance of these differing groups, then it may be said with some certainty that the simulation possesses construct validity.

It is assumed that there is a significant difference between the groups tested. This will be determined by comparing the performance of the student and primary care physician groups on the written cognitive instrument. It is assumed that the group of cardiologists is significantly different than the above two groups. If significance is not found, then validating the simulation by this means is not appropraite.

The content validity of this simulation will be determined by an expert team of cardiologists.

Scoring of Computer Based Evaluative Simulations

One of the most crucial and fundamental problems of any evaluative simulation is scoring. All the external validating instruments that were discussed rely on a reliable scoring mechanism. No matter how well these instruments are constructed, and are psychometrically sound, if the scoring of the simulation is not both reliable and realistic the validity can not be determined. Nor, does it matter how much face validity the simulation possesses if criterion related validity can not be shown. Critical to proving validity of this simulation is the development of a sound scoring algorithm.

It is crucial that the scoring algorithm realistically represent the clinical ramifications of treatment. The purpose of scoring the simulation at each data option point is to obtain an overall numeric score on the simulation for each doctor and also obtain a series of subscores on each section of the program.

This simulation contains a multitude of major data option points all of which will never be encountered in any one patient. A major data option is defined as a treatment list which is encounted when the doctor opts for medical or surgical intervention, or, as a management list which usally contains a list of options allowing the doctor to perform a physical, take a history, order laboratory tests, and/or begin medical intervention.

The scoring will be accomplished in two independent phases. The first phase will be to have a number of cardiologists take the simulation. The individual physician paths will be monitored and placed in a file. These files will consist of individual physician paths and a cumulative frequency record of all the data options that were choosen.

Once these physicians have gone through the simulation, their paths will be reconstructed on paper in the form of decision tables. Since there is no one optimal path through the simulation, the cumulative paths determined by these cardiologists results will serve as the standard "optimal" paths. Once these standards are established, a weighting scale can be instituted. It is important to note that for any major data option there exist a multiple number of routes that could be taken to reach that option point. In other words, it is possible to reach a data option by one route and choose drug 'A' for treatment and receive the maximum weighted score. It is also possible to reach that same data option and choosing the exact same drug, by taking a different and less desirable route and not receive the maximum weighted score. This indicates the need to standardize optimal pathways first.

There are hundreds of treatment and management lists incorported within this simulation. It would be almost impossible for a group of experts to weight (score) every option within every data list. It is because of this reality that the decision was made to limit and identify only the major data option points throughout the simulation. These data options constitute critically important points for diagnosis, treatment and mangement. It is these major data options that will be weighted.

A decision table will be constructed for each major data option point and all possible routes that could have been taken to reach that particular option will be specified. A 10 point weighting scale will be used to judge the plausibility of each dat majora option within the simulation. A team of experts will be employed to rate EACH major data option on a +5 to -5 scale {+5 critical (essential to do) to -5 very harmful (essential to avoid)]. From the predetermined standard the +5 weighted score will be cited for each specific option on the optimal path. It will be the job of this team to weigh the other options for each major data option list. It is quite possible and probable for any one option within a data option list to have differing weights due to different paths.

While the scoring is considerably complex, it is essential that it be done. The team of experts will work independently rating each option of each major data option list; the results of the ratings will be compiled, and the final ratings for each option will be determined. If the ratings are not consistent on any one item, the highest and lowest ratings will be discarded; and the remaining scores will be averaged to determine a final weight for the option.

Reliability of Computer Based Simulations

The last measurement consideration is that of assessing the reliability of this simulation. The ideal way of measuring reliability is to construct and develop two parallel evaluative simulations and look at individual performance on these parallel forms. Not only is it extremely differcult to develop two parallel simulations measuring the same knowledge and skill levels, but, this form of reliability estimation falls victim to all the intervening and confounding variables of any parallel forms situation. This procedure is also extremely costly and time consuming, considering that it took one and a half years to develop this present simulation.

Another, more viable procedure for establishing reliability is using a test-retest design. While this is certainly possible, this procedure is suseptable to all the intervening and confounding variables that this estimation presents.

Internal reliability estimates such as Crobach's Coefficient Alpha or KR20 all make the fundamental assumption that the individual test items are discrete and identifiable. That is, a 50 item exam has 50 dicrete, individual test items. This is not the case in an evaluative simulation. There are no test items that can be individually identified.

McGuire and Babbot (1967) in their article entitled "Simulation techniques in the measurement of problem solving skills" (1) have looked at this issue of reliability. They state that there are four characteristics that have the effect of reducing reliability as it is commonly estimated: 1) the data options are differentially weighted 2) data options are interdependent 3) there are differential amounts of feedback obtained by the examinees which results in dissimilarity among the students with respect to the nature of the problem posed by any given data option and 4) an individual student can be denied the opportunity of responding to many of the data options because of the particular decisions that s/he opts to make (i. e., may follow many different alternative paths).

It is because of these four factors that reliability estimation is extremely differcult. McGuire and Babbot take the stance that reliablity of a measure can be viewed as an attribute related to it's generalizability with respect to different universes. Of interest are the estimates of the extent to which a particular set of scores on the simulation is generalizable to many possible similar tests. Here they are referring to the characteristic that Cattell has included in his concept of "consistency of measurement" and which he specifically refers to as "homogeneity". They recommend using any of the three following techniques which have yielded essentailly similar results from their own studies; Angoff formula 12, Cronbach's Coefficient Alpha, and principal component analysis. All will estimate the consistency of measurement or generalizability of results across tests.

While it is possible to obtain a reliability estimate according to McGuire and Babbot, the four characteristics inherent in a simulation that reduce reliability have yet to be resolved. It seems possible that a coefficient can be derived which is sensitive to the multitude of confounding variables that exist in this technique. What is needed is more theoritical research in this area by those familiar with this technique and the areas of reliability estimation.

Summary

In summation, evaluative computer based simulations are both exciting and challanging measurement instruments for any eduaction environment. This mode of evaluation offers an innovative means of assessing a learners' comprehension, knowledge, and skill levels within any subject area. While there are many unresolved measurement issues, the foundation is laid for researchers to explore these vital area of interest.

Footnotes

(1) McGuire,C.H.,& Babbot,D. Simulation techniques in the measurement of problem solving skills. *Journal of Educational Measurement*,1967,*4*,1-10.

The Formative Evaluation of Bilingual Television Programs: Some Results and Suggestions for Improvement

Richard F. Lewis, Ph.D.
Research Associate
Atlantic Institute of Education

January, 1979

Paper presented at the Association for Educational Communication and Technology meeting New Orleans, La. March, 1979

ABSTRACT

The paper presents the results of two formative evaluation studies of second language programs in Nova Scotia. Both studies included procedures to determine the effectiveness of the series in achieving their objectives as well as procedures which could yield suggestions for improvement in the programs.

A second aspect of the paper concerns the improvement of formative evaluation procedures for television and other mediated programs. The suggested changes are divided into factors which can be examined independently, factors which must be examined in context and factors which must be examined in a field tryout.

The educational media have a definite role to play in helping children and adults to learn other languages. Radio, audio-tapes, films and slide sound packages and television have all been used to achieve various learning objectives. Many of the early language media presentations consisted of stand up teaching using blackboards and perhaps a few visuals but little else. However, with the maturation of educational television, more sophisticated presentations are being designed to teach languages.

Many agencies have produced series for use in second language teaching. The Ontario Educational Communications Authority in Toronto, has produced several series for use in teaching French while the *Carrascolendas* series has been used to teach Spanish in the United States. Nova Scotia School Television also produced several series to supplement teaching of French as a second language.

Most of the recent television series intended for second language learners have been subjected to formative evaluation or at least summative evaluation procedures. These studies have dealt with various aspects of the production and utilization of second language television materials. (See Laosa, 1975; Williams, Van Wart and Stanford, 1973; Fleming, 1972)

This paper deals with the improvement of formative evaluation procedures for second language television materials. The recommended procedures are based on a review of the literature and on two studies of second language teaching materials produced by Nova Scotia School Television (Lewis and Fisher, 1977; and Lewis, 1979). Although the paper deals with procedures for television programs, the comments could be applied to other media and other programs.

The Nova Scotia Studies

The recommendations on improving procedures are based on two studies of second language television programs. To provide the reader with this background, these studies are briefly summarized below.

Vive La Compagnie

Vive La Compagnie was a second language French series designed for grade eight students. In each program, three actors (and other supporting actors) participated in real-life experiences which might be faced by students. The programs had a controlled vocabulary and relied on humor and visuals to convey meaning.

The Evaluation

The evaluation was commissioned by the Nova Scotia Department of Education to determine the effectiveness of the program series achieving its objectives, which dealt with vocabulary acquisition, enjoyment and understanding of the plots. In addition to determining effectiveness, the study had to provide specific information on recommended production changes, distribution patterns, usage guidelines and supplementary materials.

Specific Questions

1. Would the students watch the programs; i.e. would their eyes be on the screen?
2. Would the students understand the words to be used in each program?
3. After watching the programs, would students understand words which were not in their repertoire before the viewing?
4. Would students understand the plots?
5. Would students enjoy the programs?

6. Would the number of programs watched result in differences in vocabulary acquisition, understanding and enjoyment?

7. How would students react to selected variables such as the plot, the characters and production techniques?

8. Would the students' attitudes to learning French and towards French people be altered by exposure to the program?

9. How would the teachers feel about the program with regard to its level, its achievement of objectives, etc.

Data Sources

The evaluation study used three basic sources of data: expert opinion, student opinion and test results, and teacher opinion. The expert opinion consisted of a contextual evaluation in which the programs were scrutinized to determine the production variables which could be improved. The actions of the characters, the pronunciation and the special effects all came under close scrutiny for suitability. In addition, the relationship of the production format, etc., to existing research recommendations for children's programs was investigated. The researcher determined whether the recommendations of experts in children's programming were being considered in the Series.

Teacher attitude was obtained using a Likert type scale similar to the students' scale. In addition, each participating teacher was interviewed to determine his/her reaction to every aspect of the program which could, if necessary, be changed or altered.

Design, Subjects, Procedure

Two hundred and eighty-eight grade eight students from 12 classes participated in the study. These students came from four different schools.

Classes were chosen at random from all the grade eight French classes in one Nova Scotia county. To determine the effects of watching more programs, some classes watched eight programs, some watched four and some watched one program; other classes did not watch any programs but participated in the study so that valid comparisons could be made. All classes completed a baseline test battery before the study began while one-half of the sample completed a pre-test. Classes viewing the program watched it on videotape in their classrooms during the regular French class. Discussion of a program followed its showing. After watching the appropriate number of programs, all the classes were tested on the same program using the techniques mentioned above. To obtain the most representative sample of teacher opinion regarding the program, any teacher using the program outside the study as well as several other groups of teachers completed the teacher questionnaire on the series.

Pre-Treatment Tests for Differences

In this study, classes had to be randomized, not students. As a result, differences between classes had to be measured before the treatment was presented. Two tests were administered to the sample prior to the study.

The Ici on Parle Test, an audio-pictorial vocabulary test, discovered significant differences between groups. In every case, means for the TV group appeared to be higher than means for the non-TV group.

Significant differences between groups were also found in responses to the questionnaire on attitude to learning French. This test comprised two sub-tests, Desire to Learn French and Motivational Intensity.

Results*

Attention

The programs maintained the attention of students watching about 80% of the time. However, certain sections of programs did seem to encourage students to stop watching programs. In addition, it appeared that students who did not understand the action (determined by self-report) stopped watching to catch up on their sleep, homework, etc. Sections of programs which were accompanied by lapses in attention were noted for further analysis.

Vocabulary Level Before Viewing

The 144 students who completed the pre-test knew 14 of the 20 words on the test. This data indicated that students could understand the major words and sentence expression used in the program. This finding showed that the assessments by the teachers and of the producers regarding vocabulary level were reasonably accurate.

Vocabulary Level After the Program

The difference between pre-test and post-test vocabulary scores indicated that students did not learn new words or phrases as a result of the program. Neither the pre-test group, nor the non pre-test group who watched the television showed a significant difference on the post-test from the non-television groups. However, it should be pointed out that the program was not designed to teach vocabulary.

* This section contains only summary information since the paper is concerned mainly with the improvement of procedures. Complete results are presented in Lewis and Fisher (1977).

The data did indicate that a student who had been exposed to a pre-test would score significantly higher than a student who was not exposed to the pre-test. As a result, one of the recommendations suggested that as an instructional device, students be exposed to the vocabulary of the program before it was viewed.

Comprehension of Plots

Students who watched more programs understood more of the action than students who watched fewer programs. It is obvious that comprehension improved as the students became accustomed to the accent and pronunciation of the characters in the Series. These findings clearly demonstrated that viewing the whole series would be more beneficial than the viewing of just one program, or even just a part of the series.

Opinion of the Program

No significant differences were recorded between students who watched one, four, or eight programs. In ranking the programs, students chose programs which they said were easy to understand, funny, presented pictures to explain the words and had a gool plot. More information on these factors was sought and obtained in the group and personal interviews after each viewing and at the end of the study. The findings of the interviews and the summary of the questionnaires provided the basis for many of the production changes which were suggested.

Opinions of Teachers

Most teachers indicated that students would like the program but that a considerable amount of preparatory and follow-up work would be needed. Most teachers felt that the students would not understand the vocabulary presented--an opinion which was completely contradicted by the actual performance of the students. A strong feeling of the teachers was that the pronunciation of one of the characters would inhibit learning the correct pronunciation. As a result, a change in characters or an improvement in pronunciation were suggested recommendations.

Attitude

The viewing of the program made no difference in the attitude scores of students. Those who watched did not attain higher attitude scores.

Problems with the Study

Experimenter Effect

In this study, the experimenter supervised the viewing of all programs and administered all tests. This procedure could hardly be called a natural situation for the classes involved. As a result, the generalizability of the data to a teacher administered situation could be called into question.

Vocabulary Testing

The same vocabulary test was used as the pre- and the post-test. The data indicated that the administration of the pre-test affected the post-test scores, i.e., students who took the pre-test socred significantly higher than students who did not complete the pre-test.

Attitude

Attitude is difficulty to measure. However, with the efforts of Gardner and Smythe (1975) attitude towards learning French can be measured. However, Gardner, Ginsberg and Smythe (1976) noted that attitudes of post-tests seem to be lower than attitudes of pre-tests where there is no intervention (such as a television program). They attribute this finding to the fact that subjects tested are alerted to attitude which they did not have to express before.

Therefore, the measurement procedures used in this study may not be as valid as others which could be designed. The deletion of a pre-test of attitude with only post-test being administered may be a more desirable procedure. Alternatively, two equivalent forms of one instrument or two highly correlated instruments could be used.

Measurement of Attention

The attention measures were subject to a great deal of error since it was impossible to be certain whether a student was attending the program or just had his head turned in the direction of the television receiver.

ALLONS-Y

The Series consisted of ten fifteen minute programs. Each program contained a number of short segments which offered a range of topics. The range of segments included appearances by a teenaged host and hostess, singers, a band, game shows, activity programs and on-site visits to locations like Chéticamp, a fishing village in Cape Breton.

Supplementary Materials

Supplementary materials included a teacher's guide and a set of flashcards depicting key vocabulary and concepts used in the programs.

The teacher's guide included a list of the program's objectives, information on the segments presented in each program, a description of the skit in each program and selected vocabulary and idiomatic expression. The guide, however, did not include any suggestions or directions for using the program.

Sixty-seven flashcards depicting vocabulary and concepts were available free of charge to teachers requesting them. These flashcards were printed in black and white on heavy cardboard.

Field Testing by Department Personnel

Three of the programs were viewed by students in 20 schools during visits by a staff member from the Education Resource Services. After the students viewed the program, they were asked to comment on the programs on an open ended questionnaire. Some of these comments are presented in a paper by Johnston-Doyle and Lewis (1978).

Purpose of the Evaluation

The purpose of the evaluation was to determine the effectiveness of the series in achieving its objectives and to make concrete recommendations for changes in the production and utilizations of the programs.

Specific Questions

1. Would viewing the program have an effect on students' ability to comprehend words and sentences?
2. Would the program help students recognize correct constructions?
3. Would the students understand the content presented in the segments?
4. Would the students identify with the teen-aged host and hostess in the program?
5. Would the students like the musical segments presented in the program?

6. Would attitude to learning French be affected by viewing the program?

7. Would attitude to the French course be affected by viewing the program?

8. Would the addition of a teacher's preparatory and follow-up work affect scores on comprehension, ability to recognize constructions, attitude to learning French and attitude to the French class?

9. What changes in the program could be recommended as a result of the testing and the opinion of teacher and students.

The study was divided into four sub-studies in order to answer the variety of questions posed by the clients. The first study concerned itself with the effect of viewing the programs without any teacher input. The second study monitored the effects of the television program and teacher input. The third study dealt with the difference between the two groups in the first two studies. The final study surveyed attitudes towards the programs.

Sample

The studies used populations of grade eight students in Nova Scotia. In studies one and two, an experimental group/control group (Viewing/non-viewing) design was used with the same teachers teaching both a viewing group and a non-viewing group. In study one, the experimental group (78 students) watched the program without teacher input while the control group (92 students) continued with the specified curriculum. In study two, the 52 viewing students watched the program and participated in teacher preparatory and follow-up activity, while the 20 control group students continued with the curriculum. In study three, the 52 students who viewed with the teacher were compared to the 78 students who did not have teacher input. Study four, which sought to measure attitudes and opinions, included 177 students who had viewed the programs.

Procedure

All students completed a baseline battery which included a test of French comprehension and vocabulary, an attitude test and a basic information questionnaire. Then, viewing groups watched eight programs while non-viewing groups proceeded with their regular work. After viewing the programs, all students completed the post-test battery. Viewing students completed a vocabulary test, a constructions test, a sentence comprehension test, a recall of skit test , three attitude tests and the tests on the opinion of each segment of the program. Non-viewing students completed only the vocabulary, sentence comprehension, constructions and three attitude tests.

Analysis

The analysis of covariance procedure developed by the Division of Educational Research Services at the University of Alberta was used to process the data. The analysis is computed on the basis of a pooled regression equation, homogeneity of regression is assumed. This procedure was used for studies 1, 2, and 3.

Results

Study 1

The purpose of study 1 was to determine whether exposure to the Series would result in an increase in post-test scores on ability and attitude.

Baseline Measures

The two groups (non-viewing and viewing) were not significantly different on the baseline measures. (See table 1). On the Ici on Parle Test, a test of French ability, the viewing group scored 25.31 while the control group scored 23.65. Both groups scored between 60 and 70 percent on the test. On the attitude test the viewing group scored 60.24 while the non-viewing group scored 56.53.

Post-test Measures

The scores of the two groups were significantly different only on the vocabulary test, after adjustment for the effect of the covariates measured in the baseline test. (See table 1). The groups were not significantly different on the sentence comprehension test, the construction test and the three attitude measures.

Table 1

Unadjusted, adjusted means and F. Ratios
for the viewing and non-viewing groups on all tests

	Unadjusted Means				Adjusted Means			
Test	Viewing	Non-Viewing	F. Ratio	df	Viewing	Non-Viewing	F. Ratio	df
Ici on Parle Test (Max=40)	25.31	23.65	2.62	1/168				
Attitude Test 1 (Max=90)	60.24	56.53	3.07	1/168				
Sentence Comprehension Test (Max=12)	4.74	4.18	3.84*	1/168	4.57	4.33	.99	1/166
Vocabulary Test (Max=16)	5.97	4.43	8.26*	1/168	5.59	4.76	5.20*	1/166
Constructions Test (Max=14)	6.06	6.20	.15	1/168	5.86	6.38	2.80	1/166
Attitude Test 2 (Max=61)	42.49	38.75	6.87*	1/168	41.66	39.44	3.21	1/166
Attitude Test 3 (Max=155)	90.17	87.18	1.37	1/168	90.46	86.94	1.85	1/166
Attitude Test 4 (Max=180)	104.18	97.46	4.05*	1/168	104.09	97.53	3.74	1/166

* p>.05

Study 2

The purpose of study 2 was to determine whether exposure to the Series would result in an increase in post-test scores on ability and attitude. However, in this study, the teacher actively worked with the class before and after the viewing, reinforcing the content.

Baseline Batteries

The viewing group scored significantly higher than the non-viewing group on the Ici on Parle test. (See table 2). The mean of the viewing group (31.85) was four points higher than that of the non-viewing group (27.85).

Post-tests

The scores of the two groups were significantly different on two tests, the vocabulary test and the sentence comprehension test (See table 2). The scores on the other tests were not significantly different, after adjusting for the covariates.

Table 2

Unadjusted, adjusted means and F. Ratios
for the Viewing and non-viewing groups on all tests

	Unadjusted Means				Adjusted Means			
Test	Viewing	Non-Viewing	F. Ratio	df	Viewing	Non-Viewing	F. Ratio	df
Ici on Parle Test (Max=40)	31.85	27.85	14.53*	1/70				
Attitude Test 1 (Max=90)	63.17	65.00	.28	1/70				
Sentence Comprehension Test (Max=12)	6.00	3.55	37.71*	1/70	5.87	3.88	24.85*	1/68
Vocabulary Test (Max=16)	11.77	4.10	81.79*	1/70	11.49	4.83	61.62*	1/68
Constructions Test (Max=14)	8.13	6.90	4.53*	1/70	7.93	7.44	.76	1/68
Attitude Test 2 (Max=61)	45.29	45.10	.00	1/70	45.69	44.07	.72	1/68
Attitude Test 3 (Max=155)	89.10	89.60	.06	1/70	88.96	89.93	.20	1/68
Attitude Test 4 (Max=180)	104.45	103.50	.14	1/70	104.63	103.26	.19	1/68

* p>.05

Study 3

The purpose of study 3 was to determine the effect of a teacher on the post-test scores. In the study, one group (TV) watched the program without any follow-up or preparatory activity while the other group (TVT) participated in activities in addition to viewing the program.

Baseline Test

The TV and Teacher (TVT) group scored significantly higher on the Ici Vocabulary Test than the TV only (TV) group. (See table 3). However, the two groups were not significantly different on the attitude battery. At 31.85, the mean for the TVT group was higher than the mean for the TV group.

Vocabulary Test

The TVT group was significantly different from the TV group. (See table 3). The adjusted scores were 6.51 for the TV group and 7.46 for the TVT group.

The efforts of the teacher obviously resulted in the increased vocabulary level of the group who had the teacher adding her input to the experience.

Sentence Comprehension Test

The adjusted scores for the two groups were not significantly different. (See table 3). The TVT group scored at 50% (6.0) before adjustment while the TV group scored at 30.5% (4.74).

The efforts of the teacher did not appear to significantly affect the results of the sentence comprehension test.

Constructions Test

The two groups were significantly different on the constructions test. (See table 3). When the scores were adjusted, the TVT group scored 9.99 while the TV group scores 7.16. Once again the teacher preparatory and following activities appeared to have a positive effect on the ability of the group to recognize correct constructions.

Attitude Tests

On all three attitude tests, the groups did not differ significantly, (See table 3), indicating that the intervention of a teacher along with the television series did not significantly affect students' opinions.

Table 3

Adjusted and unadjusted means for
TV only and TV and Teacher Groups on all variables

	Unadjusted Means				Adjusted Means			
Test	TV Only	TV and Teacher	F. Ratio	df	TV Only	TV and Teacher	F. Ratio	df
Ici on Parle Test (Max= 40)	25.31	31.85	37.43*	1/128				
Attitude Test 1 (Max=90)	60.24	63.17	1.52	1/128				
Sentence Comprehension Test (Max=12)	4.74	6.00	14.98*	1/128	5.15	5.38	.57	1/126
Vocabulary Test (Max=16)	6.06	8.13	29.79*	1/128	6.51	7.46	6.86*	1/126
Constructions Test (Max=14)	5.97	11.77	70.39*	1/128	7.16	9.99	27.26*	1/126
Attitude Test 2 (Max=61)	42.49	45.29	3.28	1/128	43.19	44.23	.62	1/126
Attitude Test 3 (Max=155)	90.17	89.10	.85	1/128	89.79	89.67	.00	1/126
Attitude Test 4 (Max=180)	104.18	104.54	.04	1/128	103.44	105.65	1.13	1/126

* $p > .05$

Study 4

The purpose of study four was to obtain student opinions of the various production and content aspects of each of the segments. The overall findings seem to indicate that for the most part, with the exception of the Chéticamp segments, the students were able to follow the action and understand what was going on. Most of the segments were considered quite enjoyable by the students. They reacted positively to the teen-aged host and hostess but said that some of the games on the game segments were probably more suited to younger children. The segment which received the highest overall rating was the skit segments, mainly because of the humour and the quick action.

The analysis of the recall of segment answers indicated that most students caught the general drift of the action but missed important details.

IMPROVING EVALUATIONS: SOME GUIDELINES

A variety of experimental designs and procedures are needed in order to conduct a formative evaluation of a second language or indeed of any television project.

One of the most important factors is the expert evaluation in which people who qualified for the test review the program independent of students to determine the answers to questions such as the suitability of the vocabulary, the accents, the correspondence with the curriculum, the plot level, etc.

Another procedure involves the use of a few students to determine empirically the vocabulary level of the program in comparison to the level of the target audience, the attention of students in small groups and other such factors.

The field evaluation involves the actual in-class testing of the program, with teacher input and without, to determine the program's ability to achieve its intended objectives. This type of procecure was quite widely used in most evaluations.

Most evaluations which included field testing use experimental group, control group designs with or without some provision for randomization or statistical control of pre-treatment differences.

In evaluating second language materials, some or all of the factors mentioned below may have to be included. The factors are divided into three categories: factors which can be examined by viewing the program, factors which need viewers interacting with the program.

Factors which can be examined independently

1. Vocabulary level of the program

The vocabulary level of the program must be within the range of the students in order that they can benefit from the program. There are several ways of determining ideal vocabulary level. One way would be to compare the list of words used in the program with the list of words used in the instruction materials which are being used in the curriculum. Another means might be to pre-test the students on the vocabulary which is used in the program. The pre-test can take the form which is used in curriculum materials which usually consists of a series of pictures in which the student circles the correct alternative which is mentioned on an audio-tape. This type of test was used in both the "Vive La Compagnie" and the "Allons-Y" evaluations. A variation was used in the "Allons-Y" evaluation in which the student numbered one of twenty pictures with the number of the phrase or word. However, this test was a bit more difficult than the traditional test in which the student circled one of four alternatives.

2. The types of constructions used

The formats of constructions will effect comprehension of students. The formats of construction used in the program must match the formats which have been used by the students in their curriculum or which have been reinforced in the teaching. Once again, the types of constructions which the students understand can be tested or can be deduced from the program materials.

3. Accents

The accent of the actors must match those which the students are used to hearing. However, there is some discussion on this fact with one camp believing that the student must get used to hearing a variety of accents while the opposite view holds that the student must hear words with the same accent. This is proven to be an especially vexing problem in Canada where the French spoken in the various regions differs in terms of accent and also differs from French spoken in France. Similarly, the types of inflections placed on words must be similar to those heard by the students. One of the ways this task can be done is to ask expert personnel to judge the accents and inflections or to pre-test on a small number of students.

4. Image presented by the actors

In any materials which have as a goal the improvement of relationships between language groups, the image presented by the actors in the second language program must be very carefully monitored. Experts in the area must be asked to evaluate the image portrayed by the actors. If the actors or actresses exhibit a negative image to the students, this may well inhibit learning.

5. Contribution of production effects

The sets, and production techniques should be related to the objectives of the program. In addition, these factors should be similar to the kinds of television to which the students are accustomed. Makeshift sets are detected very quickly by students and cannot be used if the purpose is to present professional television. In addition, the sets should not in themselves distract from the objectives.

6. Correspondence of audio and visual message

The presentation of the new vocabulary words can and should be accompanied by their equivalent in a real referent. When the French or Spanish word for cup is used, the cup should be the most prominent object on the screen. Another method would be to use a visual cue, with perhaps one of the actors pointing to the cup, to indicate precisely which object is being referred to in the auditory message. In one of the programs, the weak correspondence between audio and visual stimuli meant that students did not clearly understand which objects an auditory message was referring to.

7. Speed of the actors' speech

The actors must speak at a pace which is correct for the students. If the pace is too quick the students will not understand the speech. However, if the speech is too slow, it seems unnatural and will distract the students in other ways.

8. Level of content

If a story is being used, the level of the plot must match the interest level of the students. If, as in "Allons-Y", a segment approach is used, the segments themselves must also match the interest of the students.

9. Complexity of plot

The plot must be complex enough to hold the students' interest and yet simple enough so that the students will understand it. Teachers and other professionals can judge the complexity of plot relative to a group, however, a better means might be to use small groups of students and pilot studies to determine the suitability of the plot to the intended audience.

10. Analysis of the program into information units

Using a procedure outlined by Friedlander (1974) the program should be broken into information units so that these units can be tested in comprehension tests or vocabulary tests later on. Analysis of information units would probably take place by teachers familiar with the program or by researchers who are conducting the evaluation.

Factors which must be examined in context

1. Correspondence with the curriculum

The television program which is used should correspond in some way to the curriculum which is being used in the schools. If the vocabulary constructions or structure of the program is contrary to that of the instructional programming used in the schools, a difficulty will emerge. Once again, this evaluation could be conducted by expert researchers or by teachers who are familiar with both the program and the instructional materials being used in the school. A subsidiary point is that the program should help fulfill some of the curriculum objectives in the curriculum. It is important that the objectives of the program be somewhat the same as the objectives of the overall curriculum.

2. How the program can be used

In many cases, the schedule of the junior high and high schools would make it impossible for the program to be used on an off-the-air basis unless cable scheduling can be arranged. The accessibility of the program to the target audience, then, must be clearly identified. In some of the Nova Scotia evaluations, it was discovered that because of a chain of circumstances, many of the program materials could simply not be used by a large part of the target audience due to the inaccessibility of equipment and the problem of rotating shcedules.

3. Preparatory and follow-up work

Through interaction with teachers the ideal means of preparatory and follow-up work can and should be specified. This preparatory and follow-up work should help to achieve the program's objectives and the objectives of the overall curriculum. One of the only ways that this effect can be tested is to contrast the use of teacher input with a comparison group who do not have teacher input such as was done in the "Allons-Y" research.

4. Teacher knowledge about using the program

Through direct contact with the teachers, the evaluation should determine whether teachers are able to use the programs or whether concrete guidelines should be provided. By having the teacher view a number of programs and asking how the programs would be used, the evaluation team should be able to determine whether concrete guidelines are needed to use the program.

5. The preparation of students

The evaluation team should try to determine whether concrete preparation in the viewing of educational television materials is necessary for students. Certainly, they need no education to watch television but they may need some help in gaining the most from an educational television learning experience.

Factors which need students viewing the program

1. Attention

Attention, interest or appeal of the program, is one of the first variables to be tested when students are involved. Attention can be measured by various means ranging from a simple count of eyes on screen, to program analyzer techniques, conjugate reinforcement (Lindsley, O. R. 1962, Ogston 1975), eye movement studies (O'Bryan, 1975), distractor analysis (Palmer, 1974), or similar techniques. Lewis (1973) reviewed various means of measuring attention to audio-visual materials.

The purpose of measuring the attention is to ensure that each segment of the program is interesting enough to be watched by the students. If the student is not watching the program, all future measurement will be in jeopardy.

Attention can be tested either in the group situation or in a viewing situation in which one student or a small group watching the program is monitored by one of the techniques mentioned above to increase the vocabulary level of students, help them hear the spoken language and help them improve their ability to construct sentence in the target language. The purpose of the on-site evaluation is to determine whether any vocabulary or correct construction has been acquired through interaction with the program.

3. Comprehension of the Program

Often in the target language, although the vocabulary is at a level which can be understood, the complexity of the plot or other factors make it hard for students to understand the exact course of action. As a result, the comprehension of the program must be measured. There are various procedures to assess comprehension, the simplest one being an interview with the student to determine what he or she remembered from the program. Other means which could be used include a question technique with multiple alternatives, or a recall test in which students are simply asked to write down what happened.

4. Attitude to program segments

The enjoyment of the various segments or total program is also one of the most important areas to measure. In the case of second language instruction although the program may fail to teach any vocabulary or construction, or even in fact may fail to be completely understood by students, they may enjoy the interaction with the program. The attitude to program segments also would measure attitude to production variables, level of plot, plot complexity, the actors and any other relevant information.

5. Attitude to production

The students who view our second language programs will have viewed a great deal of television by the time they interact with one of the programs that we produce. As a result, it is important to determine whether they consider the production level to be similar to that of the commercial programs that they watch. If the production level is seen to be inferior to the programs they watch, this will likely interfere with their learning or enjoyment of the program. The attitude to production can be obtained using open ended statements or using Likert type scales.

6. Post-program monitoring

Many of the effects of the second language program will not be evident immediately after the program is viewed. However, they may well be exhibited later on after the students have had time to digest the program and fit it in to their regular work. The purpose of post-program activity monitoring is to determine the ability of the program to produce changes in the behavior of the students. The usage of vocabulary constructions learned in the program, the repetition of the types of activities produced in the program would seem to indicate whether the program would be successful in promoting long-term change. Post-program monitoring activity would have to take place through observational procedures or other similar devices.

7. Attitude to French course

One would expect that if one group of students is using a French television series while the other group is using standard curricular materials there might be a difference in attitudes. Gardner and Smythe (1975) have suggested an attitude to French course questionnaire which could probably be used to measure attitudes to the French course.

Conclusions

The formative evaluation of bilingual television programs requires a multi-faceted research approach so that the various aspects of the programs can be examined. No one approach, such as field testing, can provide guidance to producers of television programs. A variety of methods provide answers to the many questions. Many of the techniques mentioned above have been tried and tested, both the two evaluations described above and in other formative evaluations of television programs. However, some of the other procedures have emanated from the research but have not been fully tested. There is much work to be done in this field by researchers in the field of educational technology. The best research designs and measurement techniques must be marshalled to advance the learning in this area.

REFERENCES

Fleming, W. G. Evaluation of four educational television program series produced by the Ontario Educational Communications Authority. Research Paper No. 26. Toronto: O.E.C.A. 1972.

Friedlander, D. Z. Maximizing comprehension of informational films and videotape: formative evaluation during production. Mimeo. New England Instructional Television Research Center, University of Hartford, Hartford, Connecticut. 1974.

Gardner, R. C., Ginsberg, R. E. and Smythe, P. C. Attitudes and motivation in second language learning: a social psychological perspective. The Canadian Modern Language Review. 1976 32 243ff.

Gardner, R. C. and Smythe, P. C. Second language acquisition: a social psychological approach. Research Bulletin 232, Department of Psychology, University of Western Ontario, London, Ontario. 1975.

Johnston-Doyle, K. and Lewis, R.F. A content analysis of student comments on "Le masque de Marie Louise": A program in the "Allons-Y" series. Halifax, NS: Atlantic Institute of Education R. & D. Bulletin No. 1, 1978.

Laosa, Luis, M. Formative research and evaluation in the application of television media to bilingual multicultural education. California Journal of Educational Research, 1975, 25 (5) 302-309.

Lewis, Richard F. An evaluation of "Allons-Y" Halifax, NS: Atlantic Institute of Education. Technical report, 1979.

Lewis, Richard F. The effect of stimulus color and degree of realism on attention. Syracuse University, Syracuse, New York. Unpublished doctoral dissertation, 1973.

Lewis, R. F. and Fisher, G. An evaluation of the junior high school series "Vive La Compagnie" Halifax, NS: Atlantic Institute of Education. Technical report, 1977.

Lindsley, O. R. A behavioral measure of television viewing. Journal of Advertising Research, 1962, 2 2-12.

O'Bryan, K. G. Television and eye movement research. Mimeo. Toronto: Ontario Educational Communications Authority, 1975.

Ogston, D. G. An instructional television behavioral pretest measure and its relationship to learning. Journal of Educational Technology Systems, 1975 4 (2), 125-238.

Palmer, E. L. Formative research in the production of television for children. Chapter 12 of the 73rd. Yearbook of the National Society for the Study of Education. Chicago, Illinois: National Society for the Study of Education, 1974.

Williams, F., Van Wart, G. and Stanford, M. National evaluation of a Spanish/English educational television series. Austin, Texas: Center for Communication Research, University of Texas, 1973.

EDUCATIONAL DEVELOPMENT: THE REPUBLIC OF KOREA
(1970-1979)

by

Robert M. Morgan

February, 1979

LEARNING SYSTEMS INSTITUTE
Florida State University
Tallahassee, Florida 32306

Presented at AECT,
March 6, 1979,
in New Orleans

ABSTRACT

In an effort to provide a better education for more of its citizens the Korean government has undertaken a number of steps since 1970. These included a major systems study of the nations educational programs in 1970-71. This study provided the basis for a major development effort and reform of the elementary-middle school system. In 1972 a national research capability--the Korean Educational Development Institute--was created to develop and empirically validate the proposed new system. Although there have been a number of changes from the system as it was originally conceived, KEDI has been at work building the new program for the past seven years. The latest large scale field tryouts (231,000 students in 1978) indicate that the new system--largely built on instructional system design concepts--is significantly more effective than the traditional elementary-middle school programs, and not more expensive.

EDUCATIONAL DEVELOPMENT: THE REPUBLIC OF KOREA

(1970-1979)

Robert M. Morgan

A Systems Study of Korean Education - 1970

The major aim of this study was an attempt by the Republic of Korea to find if it might be able to organize its educational resources in ways that would make its educational programs more responsive to the nation's needs and, simultaneously, function more efficiently than its present educational system. The general problem could only be addressed by answering a large number of corollary questions and the Korean government invited the Florida State University to mount a project for that purpose. In the planning phase of the project it was judged that a "systems approach" to the analysis of Korea's educational sector would be more suitable and a range of expertise and competence was identified as critical to the survey and analysis process. It was quickly apparent that the varieties of needed competencies would require an interdisciplinary team of specialists and the University assembled a study team of seven persons. The study team included (1) an economist; (2) a manpower specialist; (3) an educational administrator/manager; (4) an educational technologist; (5) a teacher training specialist; (6) a systems management specialist; and (7) a behavioral scientist. The project also utilized several American and Korean consultants as they were needed. The economist, a graduate of Seoul National University, was also a native Korean.

The study team spent three months in the Republic of Korea in 1970 gathering information and data about the educational system, the economy, the nation's needs and wants for its educational programs and the resources available for potential improvement of the system. Prior to its arrival in Korea, the study team had identified several areas in which information was needed and the Central Educational Research Institute (CERI) in Seoul assembled an extensive inventory of historical data, research reports, planning documents and other relevant material. This preliminary work by CERI proved to be a great timesaver and added significantly to the efficiency of the study team. The early portion of the team's stay in Korea was devoted to identifying additional information requirements and more carefully formulating the questions that needed answers. CERI and the Ministry of Education then worked with the team in identifying the most appropriate sources of information and answers to the questions. At times during the period in Korea the study team operated as a whole, but more often the individual members of the team worked separately with counterpart Korean specialists. Before the three months were over members of the study team had visited schools at all levels throughout Korea and had talked to hundreds of teachers, school administrators and students. The team also worked closely with several Korean government agencies, most particularly the Ministry of Education (MOE) and the Economic Planning Board (EPB). By the end of the team's stay in Korea a great deal of information had been gathered, which the team then began to sift, organize and analyze.

Clearly, the study team could not in the few months available become in-depth experts in the social, economic and educational affairs of the Republic of Korea. However, a great deal was learned. From what was learned the study team formed certain impressions, developed conclusions and made some rather sweeping recommendations for educational reform in Korea. The team recognized that there was a danger of incompleteness and even inaccuracy in so hurried an information collection, but judged the available data of sufficient validity to justify its conclusions and recommendations. There were some errors or data gaps in the study that were apparent to the experienced Korean educator when they reviewed the study. Korean government officials, educators and researchers determined that these errors did not significantly effect the conclusions and recommendations.

In the study and analysis, focus was on those issues which would help the Korean Republic provide a better, more relevant education for more Korean young people at a lower unit cost and at a total cost not greater than the nation could afford. To this end, the study team collected historical, cultural and educational data, including demographic reports, economic forecasts, manpower needs projections, educational fiscal data, current and long-range educational plans and such information as was available on educational objectives and attainment. The data was analyzed in terms of future manpower needs and educational output, estimated cost benefits, and strategies for appropriate introduction of innovation and technology into the educational system. Alternative approaches to relating resources to educational objectives and problems were examined.

Economic Factors

In the post-Korean Conflict period, the Korean economy had experienced remarkable industrial progress and growth. There seemed to be general agreement among economists, both Korean and foreign, that the economy would continue to grow in the foreseeable future, would become somewhat more stable and would lead to an improved balance of payments position. The labor force was inceasing steadily and the rate of unemployment was decreasing. All indications were that in general business would be good but in an economy as dynamic as that as Korea, economists did not appear to be able to predict very many years in the future nor with much specificity what that business would consist of.

It seemed unlikely that manpower needs forecasts that extended more than two or three years into the future would be of much validity--particularly at the technical and subprofessional levels. Yet, it was at these levels that Korea was likely to develop its greatest manpower deficiencies which would be qualitative rather than quantitative. At that time in Korea, it was not unusual for a job holder to have an education which was largely unrelated to the particular job functions performed. Furthermore, manpower waste through over qualification of incumbent job holders could be observed and was anticipated to increase. A major problem was likely to result in the future from the lack of congruence between the nation's manpower requirements and the projected supply of skilled labor, subprofessionals and technicians. It was suggested that the only long range solution to these problems was a reordering of the educational priorities in the schools of Korea.

In recent years, the burden of financing public education in Korea had increasingly moved to the federal government and away from the private citizen, though the private citizen was still a heavy contributor. The formal educational system of Korea had become, to a large extent, a publicly controlled service. It was noted that purely in terms of economic needs, the quantity of schools may be sufficient for the next few years, but the quality of human resources produced by existing middle and high schools can reach neither the level for which purpose these institutions are intended nor that required for the nation's employment needs. The study team contended that investment in education is as important to economic development as investments in physical capital. As investment in education competed with investment in physical capital in the allocation of scarce national resources, it was prudent for Korea to invest relatively more in the middle schools than in the higher levels of the educational system. Two rate of return studies were undertaken in this project, which indicated that for the high school level, the rate of return (11.2%) was not much different than that for the college and university graduates (9.5%). The rate of return for the middle school education, however, was noticieably higher (20.0%). This latter rate of return was, in fact, higher tnan that for the average capital investment in Korea at that time.

There were important implications of these economic and manpower considerations for governmental and educational decision-makers in Korea. The greatest needs for manpower in the years ahead would be at the level of the middle school graduate. The rate of return on investment was astonishingly high at this level. The educational programs through the elementary and middle school were not as appropriate to the future economic needs of Korea as they could be. Using economic criterion such as earnings, employment and maximizing economic benefits, one concluded that the expansion and improvement of the elementary and middle school programs should have been given high priority. There were also social and humane arguments to support this contention.

The Contemporary Korean School System

The educational goals that characterized the Korean elementary and middle schools appeared to be restricted to the conventional academic domain. The student learning outcomes at these levels fell almost exclusively into the informational and skill categories of educational objectives. Students were acquiring the skills of reading, writing and computation, though with variable proficiency. This system seemed to be characterized by rote memorization of classically academic subjects with the overriding objective being to prepare the student for the national competitive examinations.

These exams were used to select those students who would be permitted to enter the next level of education. The exam for entrance to the middle school had been eliminated recently, but the study team did not see evidence that this action was having any positive effects on the curriculum. The existing curriculum was not as relevant to preparing Korean children to live and prosper as adults as it could have and should have been . The study team did not attempt to specify educational objectives. This could only be done by Korean educators. However, the team felt the curriculum could be broadened to include the teaching of inquiry skills, problem solving approaches and generally attend more to process objectives--and that these should not only be learning outcomes, they could also serve as effective instructional means. A general addition to the elementary-middle school curriculum and important to the enhancement of its relevance, would be the addition of preoccupational education. It was suggested that a properly conceived preoccupational program would add to the graduate's employability, his retrainability and his occupational mobility. In other words, the products of nine years of education in Korea could grow into a valuable inventory of manpower which with limited but specific additional training could be prepared for technical and skilled occupations as these needs develop and change.

The study team predicted that the present Korean educational system could not in its existing form achieve these important objectives through simple expansion or minor alteration of the existing system. An additional problem was that with 6.7 million children in school in the age range of six to fourteen years, nearly a million youngsters were out of school. By 1975 there would have been 8.2 million Korean children in the eligible age range for the elementary and middle schools. There are many reasons--manpower needs, societal stability and humanitarian--that led the study team to conclude that it was essential to expand free, universal education to include the middle school level. It was noted that free education as it is understood in the United States did not exist in Korea except for a few, very poor families. Almost all families contributed directly, through purchase of textbooks, tuition and various fees, to the cost of educating their children. When these hidden contributions to school operations were added to the visible tax contribution, the per student per year cost was estimated to average 12,878 Won for the elementary and middle schools combined.[1]

A Proposed New Educational Model - 1971

In order to develop a nation of people, all of whom had been prepared for a life of fulfillment in terms of general occupational and citizenship education, the study team suggested in its report that a nine-year, free and compulsory educational program was necessary. If the public schools through the ninth year were open to all students, were free and of uniformly high quality, it was reasonable to assume that the need for private schools and out-of-school tutoring would virtually disappear. It was also reasonable to assume that funds used for these purposes would be diverted to public school support, though in the form of public school taxation.

The vocational high schools of Korea, which enrolled slightly more than half of those students permitted to enter high school following graduation from the middle schools, were not, in the judgement of the study team, effectively serving the purposes for which they were formed. Based upon assumptions about the potential for improved academic accomplishment at the elementary and middle school level, the study team recommended that responsibility for all post-ninth grade occupational training be consolidated under a single governmental agency and that this training should have been directed exclusively to preparing people for specific jobs. These job training programs would have been of variable duration depending on the training requirements, would have been operated only as long as there were known manpower needs for the jobs in question, and would be open to qualified citizens of any age level. It was suggested that the vocational high schools of Korea should become an integral part of the job training program and cease to operate in their present form.

---

[1] All Won figures in this report are based on 1969 values. The exchange rate at that time was approximately 370 Won to the U.S. dollar.

The study team recommended that the Korean colleges and universities and the academic high schools that feed them be maintained at their present rate of growth and improvement measures of an evolutionary and gradual character be undertaken in the future. These improvement measures would be of the kind ordinarily expected in the normal course of events. Based on projections of the Ministry of Science and Technology for manpower needs at the higher levels--scientists, engineers, professional managers--the Ministry of Education would need to reappraise the enrollment quotas for the various subject areas in the universities.

The highest priorities for extraordinary change and development for Korea should have been at the elementary-middle school level. It was believed that through a substantial, but feasible, effort in the development and validation of a significantly different kind of elementary-middle school that Korea could provide an educational program of demonstrably higher quality and relevance for all age-eligible Korean youngsters. Further, it was predicted that this program once developed and installed in the nation's schools would not only be cost effective, it would in fact cost less per student to operate than was presently the case.

The new school proposed by the study team involved a number of changes from the existing system. These included changing the basic instructional unit from its present class size form to a larger grouping, introducing individualized instructional concepts and associated materials, modifying the role of the teaching staff and increasing the ratio of students to teachers, and using programmed instructional television and radio.

It was proposed that the students would be organized into "instructional units" of 300 students with the average sized Korean school having three such units. Each instructional unit would become the responsiblity of a four person teaching team whose functions would be differentiated and carefully defined in terms of what each team member contributed to the learning experience of the students. This would raise the student-teacher ratio from the present 55-1 to 75-1. It was proposed that the instructional unit (with 300 students and four staff members) would have permanently assigned to it six conventional classrooms. This would make an average student-to-classroom ratio of 50-1, down from the present 66-1. In order to get the learning group into sufficient space, the Korean government would face a socially and politically difficult decision, that of moving their schools to a double shift basis. Because of the self-study nature of much of the planned educational materials, the students would have been able to do more learning in their homes or out of school. It should have been possible to shorten the time in school for students without reducing real instructional time or learning achievement.

A basic recommendation of the study team was that the Korean elementary-middle schools be moved to a system of individualized instruction. The introduction of an individualized approach should have several benefits. It was performance based, permitted students to move at their own learning rate, and placed a larger measure of responsibility on the student for self-direction of his learning experiences. It also reduced reliance on direct teacher-to-student instruction. The basic instructional resource for that portion of the curriculum to be individualized was a "student-learning unit," which was prepared in modular and overlapping form and was packaged for ease of storage and retrieval by students. These units should have been sufficiently durable to have a use-life of four to five years. The student-learning unit contained the behavioral objectives for the unit, critical instructional materials and directions to other resources not contained in the package, and formative criterion-referenced test items which permitted the student to assess his own progress through the unit. The principles of programmed instruction were employed in the development of these units even though most of the instructional materials were not programmed instruction per se.

Another feature of the proposed program was that the teaching staff should be differentiated in a manner that calls for professional staff with differentiated specialties. This provided a better means for having the full range of competencies available in the instructional unit and made it possible to allocate different responsibilities to the individual professionals making up the team. The team functions must derive from an empirical analysis of the new learning program and required special training be given to the teaching team. The teaching team would operate under the direction of a master teacher whose main job was the management of the learning environment.

The study team proposed that a national educational radio and television distribution system be developed which would continuously transmit instructional programs during the school day. It was estimated that one and one-half to two hours of television instruction would be received by each student each day, comprising about one-third of the student's instructional day.

The type of television suggested was one which couples the principles of programmed instruction and good dramatic television production to yield programs which are interesting and will teach youngsters who are widely varied in age and socio-economic background. Television sets could be made available, programming and maintenance capability existed, and a working prototype for central transmission and nationwide relay was in place. It was estimated that a functional national educational television system could be built and installed including a television set for each instructional unit of 300 students plus an inventory of replacement sets for maintenance rotation. This television system would be an integral component of the system of instructional resources and would not be an "add-on" to the existing instructional program. It would be a form of programmed instruction developed to teach specific behaviors and would call for active responses from the student. Auxillary printed materials would be developed to go with the ITV programs in which the students would write responses, solve problems and record reactions and questions. Student learning would be closely monitored and the teacher would be furnished supportive and supplementary materials to help her work individually with any students who experience difficulty or who fall behind in the televised instruction.

In the proposed Korean elementary-middle school it was anticipated that radio instruction would be used in the context of the individualized program and be one of the instructional resources to which the student was directed.

These were extensive changes recommended by the study team in the educational processes for the Korean elementary and middle schools. Deciding upon the appropriate educational goals and operationally defining them into specific instructional objectives was a task of enormous importance to the future of Korea. The kind of individualized program being recommended by the by the study team would work best if continuous progress of each student was permitted.

Vocational Education

The part of education about which Korean leaders, both governmental and educational, invariably expressed the greatest concern was vocational preparation. The study team strongly recommended the addition, at both the primary and middle school levels, of a substantial offering in pre-occupational preparation. The learning of the specific technical job skills needed in Korea's economy could then be readily acquired on a minimum training time basis. The graduates of the proposed nine year curriculum should have solid academic preparation and should also be well grounded in the general fundamentals and prerequisites to specific job training. These changes will make the vocational-technical high schools as they have traditionally operated obsolete and unnecessary.

What would be needed was a system of job training programs which would have the following characteristics:

A. The training would be exclusively related to specific jobs that would be available as students complete the training. Manpower needs forecasting would be essential for these purposes.

B. The program, growing out of short range (two years or Less) manpower needs predictions, would be of variable duration, the training being no longer than is minimally required to prepare the trainees for the available specific job. These programs would vary from four weeks to two years in length.

C. The system would provide for the start up of new training programs, with the shortest possible lead time, as needs change. By the same token, programs would be terminated as they were no longer needed.

D. These programs would be staffed with personnel who knew the job skills being taught, with much less emphasis on degrees, teacher certification or other formal educational requirements.

E. These schools would not only train middle school graduates; they would be used for retraining of adult employees as personnel needs change.

F. There would need to be a very close liaison and coordination between the appropriate governmental agencies and the private sector to maintain the optimum relationships between training output and manpower utilization.

Evaluation

To optimize the adoption, effect and continued improvement of these innovations, it would be necessary to develop an efficient appraisal and evaluation activity which would provide policy-makers and public in general with information about the achievements, and problems which resulted from these innovations. Information for this evaluative function should be derived from the performance of individual students and various groupings of students, and at the component level, it must come from the performance of the various educational components which comprise the innovative pattern being proposed. Only by covering the range of information from the student to the component would the public and educational decision-makers be able to know the effectiveness with which the educational system was performing.

The purpose of evaluation was to provide information which would permit the continuing improvement of the educational process. Its scope should be system-wide, with a comprehensive role for evaluation at each level, ranging from tracking the progress of the individual student to assessing the performance of the total system on a national scale. Evaluation should be based on systems performance with respect to the defined goals and objectives for Korean education. With education conceptualized as an input-process-output model, the data base for the evaluation system would be described in terms of the corresponding sets of data comprehensive evaluation of the developmental effort, therefore, would provide descriptive and diagnostic feedback for the planning and development, implementation, operation, and diffusion levels. Emphasis would be placed on evaluation at the student level, the component level, the school level and the system-wide or national level. The basic purpose of this evaluation would be to provide adequate, valid and reliable information on performance assessment at each level so that the highest performance possible may be reached under the existing constraints. To create such a comprehensive evaluation system required a series of activities. The first of these was a thorough review of the nature and intent of the evaluative systems currently in operation in the Korean educational system. The second was to expand the evaluation model outlined in the study team's report into a detailed operational plan which is appropriate to Korean education, to the innovations being implemented, and to the formative requirements of these. The third step was the general process of testing and validation of the evaluation model. The fourth step was widespread implementation of the evaluative procedures. This evaluation system would help bridge the transition between the current system and the installation of the proposed system and facilitate it.

Development and Implementation - 1972

Korean educational leaders concluded that the kind of elementary--middle school proposed by the study team was sensible for the nation and appeared to be a desirable and viable alternative to the present educational condition. They have had to address several other important questions. The responsible government officials have had to judge whether or not the proposals that appear to be good for Korea are also possible. Are the various resources--people, money and time--available? What are these resources? How long would it take to develop such a system? How much would it cost to develop; how much to install nationwide; then how much to operate on a yearly basis?

The study team was optimistic that the key resource was available. This resource was a group of aggressive, technically sophisticated Korean educational researchers who were prepared to spearhead the effort. Additional support and technical staff needed to be prepared and this appeared to be feasible. These resource personnel needed an organization, under the Minister of Education, which would be funded and mandated to undertake the development and validation of the new system. The study team proposed an organization which it labeled the Korean Educational Development Institute (KEDI) whose responsibilities should include the design and tryout of the system and its components. With personnel from the MOE and representative Korean educators KEDI should reappraise the educational goals and objectives for the elementary-middle schools. It should develop definitions of desired learning outcomes at the various levels and then design and build the instructional programs to achieve these outcomes. These instructional resources--student learning units, ITV, radio and teacher directed activities--should be chosen in terms of their appropriateness to particular content and objectives in the curriculum. KEDI should be responsible for empirically demonstrating the instructional effectiveness of these new programs of learning, and should plan and develop a comprehensive educational evaluation system. The evaluation system should provide for assigning responsibility for student learning to the principal elements in the educational program, should provide for periodic audit of performance, and should permit system accountability.

The study team recommended that a three phase effort be undertaken: (1) development; (2) tryout and revision in a pilot community; and (3) nationwide dissemination. It was imagined that KEDI could best achieve the first two phases as well as continued improvement of the program through research, while the operating bureaus of MOE were best equipped to take responsibility for the third phase.

The recommendation was made to build and try out the new educational system in a single province but the developers should keep in mind that eventually the system would be used nationwide and they were designing to that end. Estimates were that it would take approximately four to five years to build and install the new system in a pilot province. Nationwide diffusion during the development phase could probably be accomplished in the same length of time, if the decision to go nationwide had been made at the onset of the program. However, by deciding to take the program nationwide only after the pilot system became operational (as the study team recommended), then an additional one to two years were required for national diffusion. The cost of development and installation on a national scale was estimated to be approximately $17,000,000.00, while the same program installed only in a single test province was estimated to cost approximately $7,500,000.

The per student annual educational cost for the new system was estimated to be 9,819 Won. This would require an annual educational budget for the Korean elementary-middle schools of 80.5 billion Won if the 8.2 million age-eligible Korean children were all enrolled in the public schools in 1975. This is contrasted with the 104.9 billion Won required for the same number of children at the 1969 per student expenditure rate.

The Korean Education Reform 1971 - 1977

The study team report with the recommendations summarized in the previous section was delivered to the Korean government in February of 1971, and seven years have elapsed since then. Although work on the proposed educational reform is still in progress enough has been accomplished to justify a summary of the activities to date.

The report was immediately submitted by the Korean government to intensive and critical review by key staff members of the Ministry of Education, the Economic Planning Board and educational leaders outside the government. Robert Morgan, the FSU study team leader, returned to Korea in April, 1971, to participate in this review process. As might be expected from such an exhaustive analysis a number of errors made by the study team were revealed and some conditions were noted to which the team had not given appropriate weight. However, these problems were not thought to be sufficient to invalidate the essential recommendations of the report. Following the reviews, with reactions both pro and con, the Korean government decided to undertake the reform program.

In terms of the Korean educational program the report included two broad targets of reform: the elementary-middle school program (the grades one through nine), and the post-ninth grade vocational educational program.

The decision was made to undertake the revision of the elementary-middle school programs at once deferring until later, action on the vocational educational programs. Clearly, this was a step of enormous importance to the Republic of Korea and one which would engage them for several years in a highly complex developmental effort. It was anticipated that the development and tryout of the new elementary-middle school curriculum would require six to seven years before it would be ready for implementation nationwide. During this projected time period a great many things would have to be accomplished. A sophisticated educational development resource would have to be chartered, staffed, housed and funded. Money for the project was to be secured through an international loan agreement but it was expected that the loan would take nearly a calendar year to negotiate. If the project was not to be delayed for a full year while waiting for loan funds the Korean Government would need to provide interim financing for project start-up.

The first concrete action which followed immediately the public announcement of the planned reform, was the creation by the Minister of Education, Dr. Kwan Shik Min, of a small nucleus group to proceed with the detailed planning efforts. Professor Yung Dug Lee, of Seoul National University, was designated the chairman of this task force and was temporarily detached from the University to the Ministry of Education. Florida State University was asked by the Korean Government to loan Morgan to the MOE from June, 1971 to September 1972 to work with the task force.

In Figure No.1 is shown a flow chart developed in 1971 by the Task Force identifying the major events in sequence which needed to be accomplished in the first eighteen months of the project.

Temporary space and funding was allocated to the Task Force and in a short time it grew to six senior professionals recruited from Korean Universities and research institutes. Wtih a small support staff the Task Force began its work. The FSU report was a general statement of Korean educational problems and relatively broad-brush descriptions of possible responses to these problems.

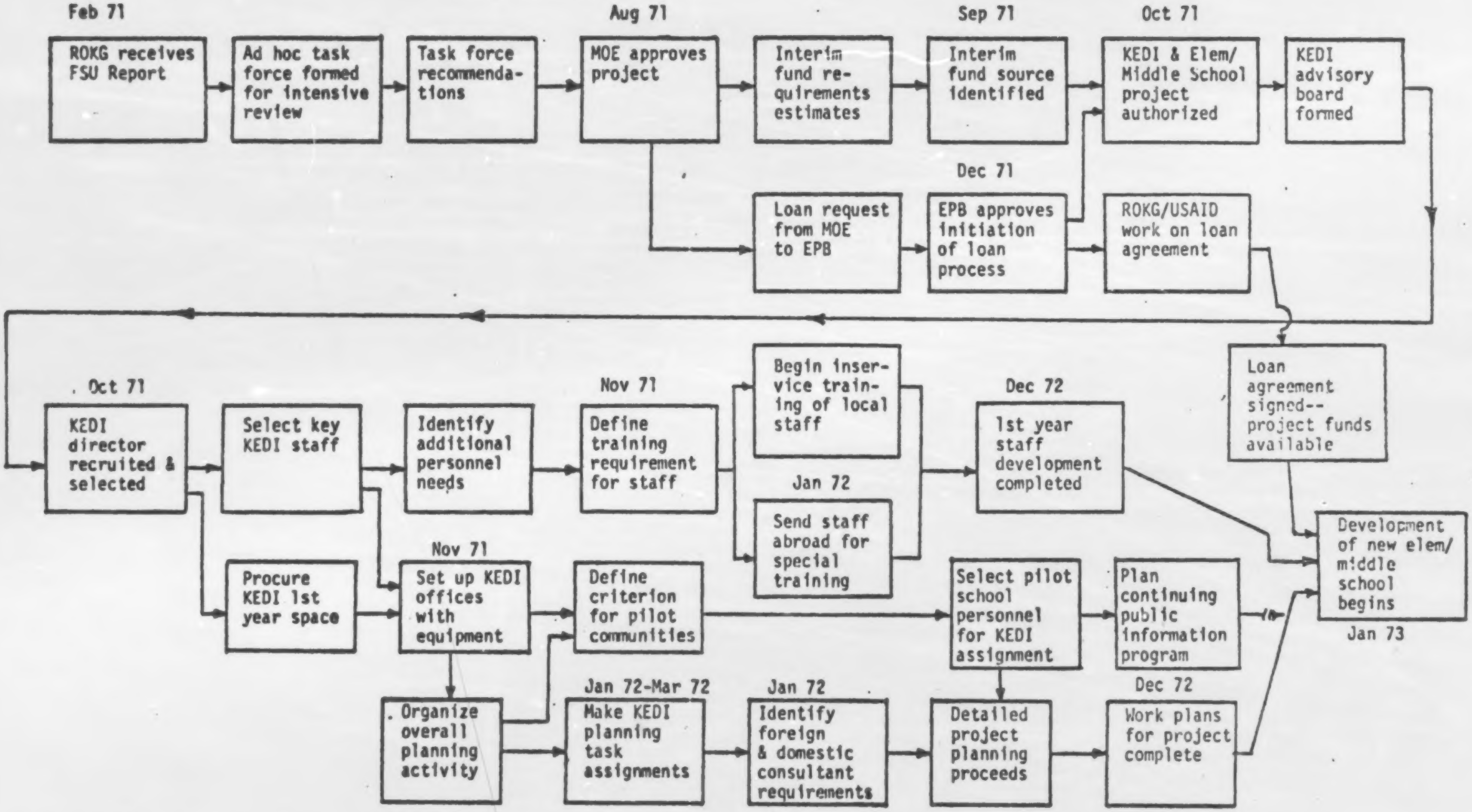

Feb 71
ROKG receives FSU Report
Ad hoc task force formed for intensive review
Task force recommendations
Aug 71
MOE approves project
Interim fund requirements estimates
Sep 71
Interim fund source identified
Oct 71
KEDI & Elem/Middle School project authorized
KEDI advisory board formed
Loan request from MOE to EPB
Dec 71
EPB approves initiation of loan process
ROKG/USAID work on loan agreement
Loan agreement signed--project funds available
Oct 71
KEDI director recruited & selected
Select key KEDI staff
Identify additional personnel needs
Nov 71
Define training requirement for staff
Begin inservice training of local staff
Jan 72
Send staff abroad for special training
Dec 72
1st year staff development completed
Development of new elem/middle school begins
Jan 73
Procure KEDI 1st year space
Nov 71
Set up KEDI offices with equipment
Define criterion for pilot communities
Select pilot school personnel for KEDI assignment
Plan continuing public information program
Organize overall planning activity
Jan 72-Mar 72
Make KEDI planning task assignments
Jan 72
Identify foreign & domestic consultant requirements
Detailed project planning proceeds
Dec 72
Work plans for project complete

Figure No. 1
Activity Flow Sequence for Elementary-Middle School
Educational Development Project (Feb, 1971 - January, 1973)

It was an impetus for reform not a detailed blueprint for action.

The first priority the Korean task force set for itself was the development of a complete plan for the management of the reform project. They estimated the number of instructional hours for which materials of various kinds would need to be developed. For all the subject areas for the nine grade levels teaching materials for nearly 8,000 hours of instruction would be required. This included instructional television, the multi-media mastery learning units, texts and workbooks, and teacher guides. Development groups in other Korean research institutes had some experience in these kind of activities and they assisted in the calculation of the numbers of people and the different specialities that would be required to complete the work in the alloted time period. From this analysis it was estimated that the development group would need to consist of approximately 300 people--professional, technical and support personnel. Among this total complement of manpower a wide range of competencies were represented. Writers, editors, photographers, ITV producers and technicians, instructional designers were only a few of the critical for which experienced people had to be recruited or new people had to be trained.

An outgrowth of this effort was the beginning definition of the organizational structure of the planned Korean Educational Development Institute which had been described in the FSU recommendations. Other by-products included early projections of office, laboratory, studio and other space requirements, as well as the needs for generic and specialized equipment. The Program Evaluation and Review Technique (PERT) was used in this early stage to show how human and material resources would need to be orchestrated in time to insure the success of the E-M project. Also revealed at this stage were the first projections of financial requirements by month and year for the life of the project. A reasonably accurate estimate of cash flow and total money needed was essential for purposes of national budgeting and was a prerequisite to the completion of the international loan agreements.

Parallel to this process of man-loading the project was a nationwide survey of human resources with requisite skills and experience which could be employed to work on the project. A comparison of the numbers and kinds of people who were available confirmed the earlier judgement by the study team that twelve to eighteen months of training time would be required to prepare a sufficient staff. There were enough people identified to begin staffing KEDI, particularly in the most senior professional, managerial positions. At this point some of these senior people were employed by the Task Force and the Task Force gradually began to function as the leadership staff of KEDI-though the Institute had not yet been incorporated by the national assembly. Late in 1971 Dr. Yung Dug Lee was selected by the Minister of Education to serve as the Director of KEDI.

During the last quarter of 1971 the KEDI staff focused on two major activities. These were: (1) an intensive series of meetings, conferences and presentations on the E-M project and KEDI, held with teachers, school administrators and boards, lay groups and representatives from the public media; and (2) the writing of the international loan agreement in cooperation with the Ministry of Finance and the Economic Planning Bureau. The first of these activities was essential to broaden the base of support for the reform effort, respond to questions or criticisms and to secure the cooperation of educators throughout the nation. Very little public disclosure of the nature of the E-M project had been made prior to this time, though the systematic publication of all aspects of the project to all levels of Korean society has been an important continuing process since the early part of 1972.

In late February, 1972, a series of events occurred which threatened to seriously jeopardize the entire project. By this time the loan agreement between the U.S. Government and the Republic of Korea was in the final stages of negotiation. The threat resulted from a small group of Koreans, who, for variety of reasons were opposed to the reform project. This group gained access to the Korean President, Chung Hee Park, and asserted that the project represented a danger for the Republic and would be less effective and more costly than the existing system. As a consequence the President ordered all project activities halted until the project could be thoroughly studied by a special Presidential Commission which he appointed. The Commission, made up of leading citizens, educators and government officials who had not been previously involved in the project, undertook an intensive review of the reform plan hearing from professional educators, government officials and any other Koreans who could contribute to their judgement. These hearings lasted for several weeks, with all points of view being heard, before the final recommendations were made to President Park. The Commission recognized that certain risks were involved--as in any major change effort--but they judged the project to be soundly conceived and that the potential benefits to Korea far outweighed any associated dangers. The various allegations about the reform plan were judged to be unimportant or without substance and the President was advised to let the work proceed--which he did. A salutory side-effect of this interesting episode was that many key people, including the President and the Prime Minister, had achieved a thorough understanding in all that was involved in the proposed reform. This depth of understanding by the highest level Koreans was to pay dividends for the project in the years ahead.

On July 31, 1972 the Korean National Assembly approved the international loan agreement for the educational development project and one month later, August 30, 1972 the Korean Educational Development Institute was officially registered as a special foundation. While there were still a few legal loose ends to tie up, for all intents and purposes an enduring, sanctioned organization for educational change had been created. A board of directors was designated who quickly confirmed Dr. Yung Dug Lee's appointment as KEDI's first director.

While the raison d'etre for KEDI--at least at the beginning--was the conduct of the Elementary-Middle School project it was expected that it would also serve other purposes for the Ministry of Education. The Project Loan Proposal states four broad aims of the Institute.

KEDI Purposes:

1. To determine educational ideals and objectives which reflect the cultural heritage, social reality and future direction for the Korean society.
2. To reformulate and systematize educational content to correspond to educational objectives.
3. To develop and utilize modern educational methods, facilities and materials to achieve an effective and economically efficient program of education.
4. To establish a comprehensive research and development agency to assist the Ministry of Education in formulating educational policy for the nation.

These statements indicate a considerably broader purpose for KEDI than just the E-M project and activities of the Institute since its start-up has reflected these broader purposes. KEDI has inaugurated a "High School of the Air" program which allows out of school adults to continue their high school education through radio instruction in cooperation with participating local high schools. KEDI has also been funded to develop special curriculum in the area of population education. It has and is participating in several other similar kinds of projects at the request of the MOE which are worthy contributions to educational change but are only peripherally related to the E-M project.

Still the Elementary-Middle School project is the main driving force of KEDI and much work has been completed on that project. For the first couple of years the attention of the KEDI leadership had to be focused on institution development. This took longer and was more difficult than any of the original planners had anticipated. KEDI entered into an agreement with FSU for help with their staff development and for on-call technical assistance. Clearly one of the most urgent needs of KEDI was to rapidly train and expand its professional staff.

Since 1974 the curriculum and instructional development effort has represented the main activities of KEDI on the E-M project. Principal areas of activities include:

1. Development and tryout of new instructional delivery models including classroom, media and management considerations.
2. Development of instructional materials and programs (print, ITV and radio) compatible with MOE curriculum and suitable for the new instructional delivery system.
3. Construction and operation of broadcasting and transmitting facilities for television and radio.
4. Experimentation and evaluation.
5. Development of teacher-training programs and materials; training of participating teachers and administrators.
6. Development and maintenance of a network of contacts and communication linkages at provincial and local levels.

Field Validation and National Diffusion - 1978-79

As the project has matured and the KEDI leadership has gained experience many changes in the original plan of action have been made. The systematic emphasis of the FSU study appear to have survived but many of the specific strategies for implementation have been modified--sometimes substantially. The E-M project called for several major changes in Korea's elementary and middle programs. While the strongest rationale in 1971-72 for the reform was in the cost area, this was soon reduced as a priority by KEDI and the Ministry of Education. One reason for this shift was the Korean economy and its growth during this period. The FSU report had projected a continuing shortfall in the possible school enrollments based on Korea's predictions in growth rate of its GNP from 1970 to 1978, assuming that the share of GNP allocated education would remain relatively constant. The report also anticipated a continuing shortage of teachers. Both of the predictions missed the mark. GNP grew at a much higher than predicted rate and more teachers were trained and available for classroom assignment than had been expected because of substantial increases in salaries. With a fairly large increase in per capita GNP there was simply more absolute money available to support education, even though the relative amount remained about the same.

There are two consequences of this: (1) Korea has been able to enroll nearly 90 percent of the age cohort in the first nine years of education and expect to make attendance for elementary middle school education compulsory and available for all in the early 1980s. (2) KEDI, relieved in large measure of the numbers problem, could concentrate on the E-M project in improving instructional efficiency and content relevance. With the reduced pressure to save money, some of the more politically and socially troublesome recommendations of the FSU report were abandoned early in the project. These included double-shifting of schools, differentiated teaching staff functions, and grouping students into larger learning groups. ITV and instructional radio have also experienced a major change in their respective roles in instructional delivery, but for different reasons. It was originally intended that they would be an integral part of the daily instruction of all students, and indeed, the instruction delivered by ITV and radio would not be given any other way. In other words, ITV and radio were to have been a critical part of the instructional delivery system. These elements had been included in the proposed instructional repertoire to contribute both to cost-saving and improvement in the quality of instruction.

The Korean government entered into a contract with the Tethered Communication Corporation (T-Com) of Westinghouse in December, 1972 for the purchase of a TV and radio broadcast transmission system to be installed in the north central part of the country. KEDI built a 27,050 square meter facility to house and support the T-Com transmitting operation. The T-Com transmitting system included a helium filled tethered balloon flying at an altitude of 10,000 feet to which transmitters are affixed. It was planned for KEDI's ITV and radio programs to originate at the main studios near Seoul, be relayed to the T-Com transmitters carried by the balloon on a C Band microwave link and be retransmitted to television and radio receivers in school classrooms. Two UHF channels were dedicated for ITV and one FM channel for radio. This first site was to have been one of a two phase operation. The second phase would have seen a second balloon installed in the southern part of the Republic and the two sites together would have distributed an acceptable signal to the entire country, including the off-shore islands.

This aspect of KEDI's project was fraught with technical problems from the onset, both in terms of the aerodynamic stability of the balloon and the quality of the signal. After a few trial broadcasts in late 1975, the technical problems were judged insurmountable in the Korean climate. The system was declared unworkable and the site was dismantled in 1977 (Masoner, 1979).

During the same period KEDI had completed the installation of its TV and radio production studios and many lessons using these media had been produced. KEDI made the decision to develop a groundbased microwave relay system when the T-Com technology was abandoned. Such a system would parallel the existing microwave relay system presently owned and operated by the Korean Broadcasting System (KBS) under the jurisdiction of the Ministry of Communications. In early 1979, discussion were still underway in Korea as to whether responsibility for the educational transmission network would remain with KEDI and the MOE or be assigned to KBS and the Ministry of Communication. However this decision goes, it seems certain that there will be a nationwide broadcast network dedicated solely to educational programs, and that KEDI will retain programming, production and scheduling control, even if it does not own and operate the transmitter system.

While the T-Com Corporation indemnified Korea for its financial losses in this venture, the time lost by KEDI cannot be replaced, nor can the credibility it lost with the Korean public, as a result of promising national color television broadcasts which it has been unable to deliver.

Because of the failure of the tethered balloon transmission system bought from Westinghouse it was impossible for KEDI to build a reliance on the ITV and radio components during the devleopment period. The instruction to have been provided by ITV and radio had to be given in other ways, using teachers and instructional materials which could be made available to remote schools. While a great many instructional programs in ITV and radio were developed by KEDI--and most of these were tried out in classrooms on a fragmental basis--the new instructional system neither depended upon them nor provided for their inclusion as an integral instructional element. It now appears these programs will be broadcast as adjunct to the new teaching system, rather than as an essential part of it.

What then remains of the ambitious reform project started in 1971? As it turns out, a great deal remains and probably all of the truly essential concepts have been preserved. KEDI seriously addressed the task of analyzing the existing curriculum and the educational needs of Korea's citizenry and systematically identified the places where the curriculum was not sufficiently responsive to these needs. It then developed relatively inexpensive instructional materials and student learning evaluation instruments which articulated with those text and other materials already in common use. It devised a teaching/learning process which could be managed by the existing teaching personnel and school administrators. It then devised teaching guides and intensive training programs for teachers and administrators to prepare them to function effectively as the orchestrators of the new instructional programs.

Finally, they conducted an iterative series of validation tryouts of the new instructional elements and the elements in aggregation. These field tryouts were started in May, 1973 and only 745 students from two schools in Seoul were involved. There were two subject areas--mathematics and science--in the third and fifth grades which were tested. This test, the first of several, was the means of formative evaluation and feedback of the new instruction. It told the KEDI developers what needed to be changed in the materials and processes tested and provided a basis of experience for future development. The indicated changes were on the basis of how effectively the programs caused children to learn. It was the beginning of a competency based program of student learning.

During the successive demonstration tryouts KEDI learned a great deal about areas where remediation materials were needed for some students and enrichment programs were needed for others. Student learning data suggested areas for individualized or self-teaching and pointed up those areas where teacher directed group instruction was most appropriate. Most important, the teachers and students know, on a daily basis what the desired learning outcomes were and had a means of assessing progress toward these ends. The latest tryout, completed in 1978, included 231,567 students in the first six grades of 231 schools from throughout the nation. All subject areas were included. Student achievement data, student, parent and teacher reaction, and process efficiency data were collected from both the tryout schools and a comparable sample of control schools providing conventional instruction. As of the writing of this paper the massive array of data collected during this tryout has not been completely analyzed but the results on all of the major variables significantly favor the new program. In the several tryouts--four small scale and four large--since 1973 the achievement levels have generally been higher for the demonstration students than for the comparison group. As a rule, the achievement differences between the two groups tended to become larger with each successive tryout.

In 1978, the President of the Republic appointed an external commission to conduct an independent evaluation study of the new E-M program. This group assessed student and teacher attitudes toward the new program as well as community reaction. They also selected 18 schools with some 3500 students enrolled, and directed that the new KEDI system be implemented in these schools for five months in six basic subject areas, and identified a group of traditional schools to serve as the control. They found that mean achievement across all of the subject areas was 24 percent higher in the experimental group than in the control group, and that 30 percent more of the experimental students achieved subject mastery. Another finding was that experimental students in rural schools--rural students were usually well behind urban students in achievement--performed as well as students enrolled in city schools. The commission recommended an orderly implementation of the new E-M program in all of Korea's schools.

REFERENCES

Masoner, Paul H. (Editor) Analytical Case Study of the Korean Educational Development Institute. (1979) American Association of Colleges for Teacher Education. Washington, D.C.

Morgan, Robert M. and Chadwick, C. B. Systems Analysis for Educationl Change: The Republic of Korea (1971). The University of Florida Press, Gainesville, Florida.

KEDI has produced a large number of research reports, monographs and position papers which have much potential use for professional educators. Many of these are available in English and may be obtained directly from KEDI.